Question number, 1&2

chapter:-1 (Descriptive statistics & Basic Probablity)

1.1) Introduction in statistics and its importance in engineering

* Introduction to statistics

Statistics is the collection of information in the form of data related to any concerned study. Usually the data is collected by survey method.

According to Lovitt "statistics is the science which deals with collection, classification & tabulation of numerical facts on the basis for explnation, description & comparison of phenomenon."

Where as carton and cowden defined it as "the science which deals with collection, tabulation, analysis and interpretion of numerical data.

* Importance of statistics in the field of engineering.

In addition to fields like biology, psychology, education, economics, etc. statistics have a wide range of application in the field of engineering. For example like testing of materials, system test and analysis, control of production process, etc.

Few importance of statistics in different field of engineering are listed below.

a) Statistical technique can be used to test and construct engineering systems & experiments.

b) To maintain the standards of manufacturing process and their products, statistic plays a vital role.

c) Statiscal study can be used to study the repetitive operation in order to set standard.

d) Statistic can play a vital role to check the ability, performance & reliability of any system, machine or equipment.

* Function of statistics

The main function of statistics are:-

a) To simplify complexity.
b) To prevent the fact in the definite form.
c) To determine the relationship between different phenomenon.
d) To help in forcasting.
e) To help in formulating & testing of hypothesis.
f) To draw valid inferences or conclusions.

* Limitation of statistics

There are few limitations of statistics which are,

a) Statistics is not suited for qualitative study

Statistics is a science which uses quantitative data for study purpose. For example a set of numerical data

whereas qualitative phenomenon like honesty, poverty, culture etc cannot be expressed under statistics. However ~~quantitative~~ qualitative data can be converted into quantitative data by numerical description to corresponding qualitative data.

For example,

The inteligence of group of individual can be measured by the score they obtain in certain tests.

b) Statistics doesnot study individuals.

It deals with aggregate and does not give any recognition to individual items of a series.

c) statistical laws are not exact.

Statistical laws are only approximations are not exact. On the basis of statistical analysis we can only talk in terms of probablity and chance, not in terms of certainty.

d) Statistics is liable to misused

statistical methods can be worthless tool in the hand of inexperienced & clumsy. The requirement of experience and skill for judicious use of statistical method restricts their use to only experts & hence limits the chance of mass popularity.

* Standard deviation

Standard deviation usually denoted by the Greek letter σ (sigma) & defined as the positive square root of the arithmetic mean of the square of the deviation of the given values from their arithmetic mean.

1) Individual series

$$\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

OR, $$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$$

where $d = X - A$, A = assumed mean

2) Discrete series

$$\sigma = \sqrt{\frac{\Sigma fx^2}{N} - \left(\frac{\Sigma fx}{N}\right)^2}$$

OR, $$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

where, $d = X - A$, A = assumed mean

3) Continuous series

$$\sigma = \sqrt{\frac{\Sigma fm^2}{N} - \left(\frac{\Sigma fm}{N}\right)^2}$$

OR, $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$

where, $d = X - A$, A = assumed mean.

OR, ~~spe~~ step deviation method

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times h$$

where $d' = \frac{X-A}{h}$, h = width of class interval.

* Variance

The square of standard deviation is called the variance.

2070 chaitra

Q) Calculate the standard deviation from the following data regarding marks obtained by students in a test.

Marks.	1	2	3	4	5	6	7	8	9
No. of students.	32	41	57	98	123	83	46	17	3

What will be the value of standard deviation if the marks obtained by ~~the~~ each of the students are increased by one?

Soln Given,

The table is given below,

Marks (x)	No. of students (F)	Fx	Fx^2
1	32	32	32
2	41	82	164
3	57	171	513
4	98	392	1568
5	123	615	3075
6	83	498	2988
7	46	322	2254
8	17	136	1088
9	3	27	~~333~~ 243
	N = 500	$\Sigma fx = 2275$	$\Sigma Fx^2 = \cancel{12015}$ 11925

Now, $s.d(\sigma) = \sqrt{\frac{\Sigma fx^2}{N} - \left(\frac{\Sigma fx}{N}\right)^2}$

$$= \sqrt{\frac{\overset{11925}{\cancel{12015}}}{500} - \left(\frac{2275}{500}\right)^2}$$

$\therefore \sigma = \cancel{1.82}$

Also, when marks obtained by each of the students are increased by one then,

marks	2	3	4	5	6	7	8	9	10
No. of students	32	41	57	98	123	83	46	17	3

So,

for sample,

$$\text{standard deviation}(s) = \sqrt{\frac{n\sum x^2 - (\sum x)^2}{n(n-1)}}$$

Marks(x)	No. of students (f)	fx	fx^2
2	32	64	128
3	41	123	369
4	57	228	912
5	98	490	2450
6	123	738	4428
7	83	581	4067
8	46	368	2944
9	17	153	1377
10	3	30	300
	N = 500	$\Sigma fx = 2775$	$\Sigma fx^2 = 16975$

$$\therefore \text{standard deviation}(\sigma) = \sqrt{\frac{\Sigma fx^2}{N} - \left(\frac{\Sigma fx}{N}\right)^2}$$

$$\text{or, } \sigma = \sqrt{\frac{16975}{500} - \left(\frac{2775}{500}\right)^2}$$

$$\therefore \sigma = 1.77$$

070 Ashad

8) In statistics paper five candidates obtained the marks as 33, 38, 48, 59, & 72. Calculate the mean & standard deviation of these marks. If ~~the~~ 10 marks are added for each students, what will be mean & standard deviation?

Soln Let the data given are population.

Then, mean $(\mu) = \frac{\Sigma x}{n}$ & sd $(\sigma) = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$

where $n = 5$, $\Sigma x = 250$ & $\Sigma x^2 = 13502$

so, mean $(\mu) = \frac{\Sigma x}{n} = \frac{250}{5} = \underline{\underline{50}}$

& standard deviation $(\sigma) = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$

$= \sqrt{\frac{13502}{5} - \left(\frac{250}{5}\right)^2}$

$\therefore \sigma = \underline{\underline{14.156}}$

When, 10 marks are added to each students,

then $\Sigma x = 300$, $\Sigma x^2 = 19002$

so, mean $(\mu) = \frac{\Sigma x}{n} = \frac{300}{5} = \underline{\underline{60}}$

& standard deviation $(\sigma) = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$

$= \sqrt{\frac{19002}{5} - \left(\frac{300}{5}\right)^2}$

$\therefore \sigma = \underline{\underline{14.177}}$

069 chaitra

8) The following are data on the breaking strength of 3 kinds of materi

Material 1	144	181	200	187	169	171
Material 2	186	194	176	182	133	183
material 3	197	165	180	198	175	164

i) calculate the average breaking strength & the median breaking strength for each material.
ii) Calculate the standard deviation & variance for each material.

Sol? Given, $n = 6$

	Material I		Material II		Material III	
	x	x^2	x	x^2	x	x^2
1^{st}	144	20736	5^{th} 186	34596	5^{th} 197	38809
4^{th}	181	32761	6^{th} 194	37636	2^{nd} 165	27225
6^{th}	200	40000	2^{nd} 176	30976	4^{th} 180	32400
5^{th}	187	34969 ~~28561~~	3^{rd} 182	33124	6^{th} 198	39204
2^{nd}	169	28561	1^{st} 133	17689	3^{rd} 175	30625
3^{rd}	171	29241	4^{th} 183	33489	1^{st} 164	26896
	$\Sigma x =$	$\Sigma x^2 =$	$\Sigma x =$	$\Sigma x^2 =$	$\Sigma x =$	$\Sigma x^2 =$
	1052	186268	1054	187510	1079	195159

i) Mean,

For material I, $\mu = \frac{\Sigma x}{n} = \frac{1052}{6} = 175.333$

For material II, $\mu = \frac{\Sigma x}{n} = \frac{1054}{6} = 175.667$

for material III, $\mu = \frac{\Sigma x}{n} = \frac{1079}{6} = 197.833$

Median

Here, no. of observation is even, so the median is the arithmetic mean between $(n/2)^{th}$ & $(\frac{n+2}{2})^{th}$ term.

Here, it will be in between 3^{rd} & 4^{th} item.

For material I.

$$Md = \frac{\cancel{200+187}\ 171+181}{2} = 176$$

For material II

$$Md = \frac{182+183}{2} = 182.5$$

For material III

$$Md = \frac{175+180}{2} = 177.5$$

ii) standard deviation

For material I

$$\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2} = \sqrt{\frac{186268}{6} - \left(\frac{1052}{6}\right)^2} = 17.4$$

For material II

$$\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2} = \sqrt{\frac{187510}{6} - \left(\frac{1054}{6}\right)^2} = 19.82$$

For material III

$$\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2} = \sqrt{\frac{195159}{6} - \left(\frac{1079}{6}\right)^2} = 13.66$$

ii) Variance

$$s^2 = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)}$$

For Material I

$$s^2 = \frac{6 \times 186268 - (1052)^2}{6(6-1)} = 363.467$$

$\therefore s = \underline{\underline{19.06}}$

For material II

$$s^2 = \frac{6 \times 187510 - (1054)^2}{6(6-1)} = 471.467$$

$\therefore s = \underline{\underline{21.71}}$

For material III

$$s^2 = \frac{6 \times 195159 - (1079)^2}{6(6-1)} = 223.767$$

$\therefore s = \underline{\underline{14.959}}$

068 chaitra

8) Write any four characteristics of ideal measure of central tendency. For a group of 16 candidates, the mean and standard deviation were found to be 20 and 5 respectively. Later is discovered that the score 32 was measured as 23. Find the correct mean & standard deviation.

Soln. Any four characteristics of ideal measure of central tendency are given below.

a) It shouldnot be affected by extreme value.

b) It shouldbe rigidly defined & easy to understand.

c) It should be capable of future mathematical treatment.
d) It should not be subjected to complicated & tedious calculations.

extra { e) It should be stable ~~&~~ with regard to sampling.
f) It should be band on all observations. }

Numerical part

Soln Given,

$N = 16$

mean $(\mu) = \cancel{10}\ 20$

standard deviation $(\sigma) = \cancel{8}\ 5$ ~~, $\Sigma x = N\mu = 320$~~

~~Now,~~

~~if the wrong terme is omitted then,~~

~~$N = 20 - 1 \quad 16 - 1 = 15$~~

~~sum of 15 items = 320 − 23 = 297~~

As, we know that

$$\mu\ \cancel{\bar{x}} = \frac{\Sigma x}{N}$$

So, $\Sigma x = \cancel{\Sigma x}\ \mu \cdot N$

$= 20 \times 16 = 320$

& $\Sigma x^2 = N(\mu^2 + \sigma^2)$

$= 16(20^2 + 5^2)$

$= 6800$

Now, correct $\Sigma x = 320 - 23 + 32 = 329$

correct $\Sigma x^2 = 6800 - 23^2 + 32^2 = 7295$

So,

correct mean is $\mu = \frac{\Sigma x}{N} = \frac{329}{16} = 20.5625$.

correct standard deviation, $\sigma = \sqrt{\frac{\Sigma x^2}{N} - \mu^2}$

$$= \sqrt{\frac{7295}{16} - (20.5625)^2}$$

$$= 5.755.$$

Q2 chaitra

Q) What are the differences between measures of central tendency & measure of dispersion? The mean & standard deviation of 20 items is found to be 10 & 2 respectively. At the time of checking it was found that one item 8 was incorrect. calculate the mean & standard deviation if
i) the wrong term is omitted
ii) it is replaced by 12.

Solⁿ Central tendency is the middle point of a distribution. Measures of central tendency are also called measure of location or averages. Whereas,
The measure of dispersion is the scatterness of the items from the central value. So, dispersion is defined as the measure of variation in the items from the central value.

Numerical part.

soln Given, $N = 20$

mean $(\mu) = 10$

standard deviation $(\sigma) = 2$

Since $\mu = \frac{\Sigma x}{N}$

$\therefore \Sigma x = \mu . N = 10 \times 20 = 200$

Also, $\Sigma x^2 = N(\mu^2 + \sigma^2)$

$= 20(10^2 + 2^2)$

$= 2080$

Now,

if the wrong item is omitted then,

$N = 20 - 1 = 19$

$\Sigma x = 200 - 8 = 192$

$\Sigma x^2 = 2080 - 8^2 = 2016$

$\therefore$ mean, $\mu = \frac{\Sigma x}{N} = \frac{192}{19} = 10.11$

& s.d, $\sigma = \sqrt{\frac{\Sigma x^2}{N} - \mu^2}$

$= \sqrt{\frac{2016}{19} - \left(\frac{192}{19}\right)^2}$

$= 1.998$

Then,

if wrong item is replaced by 12 then,

$N = 20$

$\Sigma x = 192 + 12 = 204$

$$\Sigma x^2 = 2016 + 12^2 = 2160$$

$$\therefore \text{mean}, \mu = \frac{\Sigma x}{N} = \frac{204}{20} = 10.2$$

$$\& \quad s.d = \sigma = \sqrt{\frac{\Sigma x^2}{N} - \left(\frac{\Sigma x}{N}\right)^2}$$

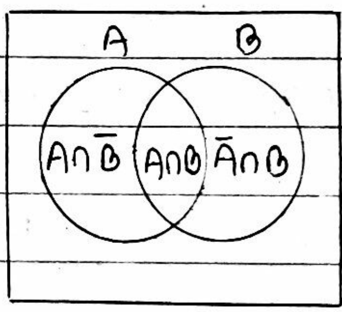
$$= \sqrt{\frac{2160}{20} - \left(\frac{204}{20}\right)^2}$$

$$= 1.99$$

1.4) Basic probablity additive law, multiplicative law, Baye's theorem.

* Addition theorem of probablity

Proof



If A & B are any two events defined in sample space and are not disjoint

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

From venn diagram,

It is clear that $A \cup B = (A \cap \bar{B}) \cup (A \cap B) \cup (\bar{A} \cap B)$

We have,

$$P(A \cup B) = P[(A \cap \bar{B}) \cup (A \cap B) \cup (\bar{A} \cap B)]$$

$$= P(A \cap \bar{B}) + P(A \cap B) + P(\bar{A} \cap B)$$

$$= P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B)$$

$$\therefore P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If A, B & C are events then the probability of occurance of at least one of three events is,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C).$$

* Multiplication law of probablity & conditional probablity

If 'A' & 'B' are two dependent event.

$$P(A \text{ and } B) = P(A \cap B) = P(A) \cdot P(B/A), \quad P(A) > 0$$
$$= P(B) \cdot P(A/B), \quad P(B) > 0$$

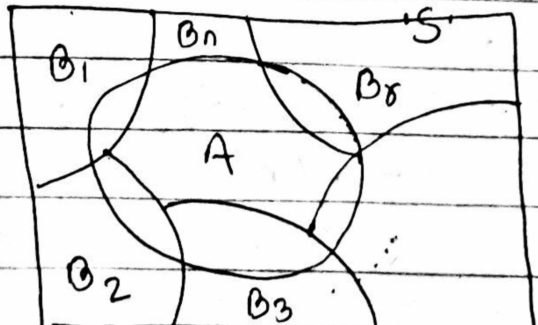
where $P(B/A)$ is a conditional probability of occurance of B when A has already happened and $P(A/B)$ is the conditional probablity of occurance of A when the event B has already happened.

1) $P(A \cap B \cap C) = P(A) \cdot P(B/A) \cdot P\left(\frac{C}{A \cap B}\right)$

$$= P(B) \cdot P(A/B) \cdot P\left(\frac{C}{A \cap B}\right)$$

$$= P(C) \cdot P(B/C) \cdot P\left(\frac{A}{B \cap C}\right)$$

* Baye's theorem

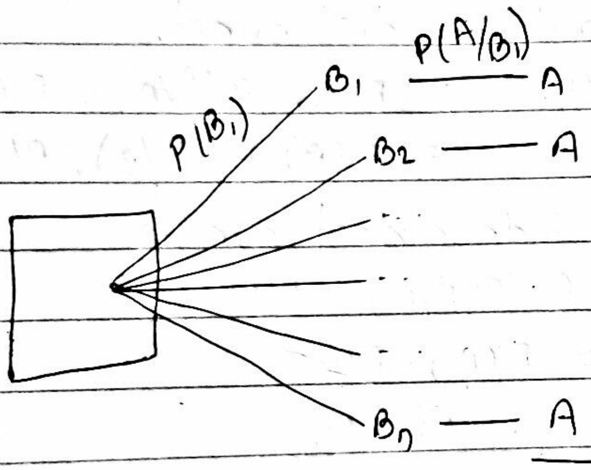


let $B_1, B_2, B_3, \ldots B_n$ be a mutually disjoint events of sampel space S & A be any event that occurs with $B_1, B_2, \ldots B_n$ then Baye's theorem states that

$$P\left(\frac{B_r}{A}\right) = \frac{P(B_r \cap A)}{P(A)} = \frac{P(B_r) \cdot P(A/B_r)}{P(A)}$$

$$= \frac{P(B_r) \cdot P(A/B_r)}{\sum_{i=1}^{n} P(B_i) \cdot P(A/B_i)}$$

where, $r = 1, 2, 3, 4 \ldots n$



$B_1 \xrightarrow{P(A/B_1)} A = P(B_1 \cap A) = P(B_1) \cdot P(A/B_1)$

$B_2 — A = P(B_2 \cap A) = P(B_2) \cdot P(A/B_2)$

$B_n — A = P(B_n \cap A) = P(B_n) \cdot P(A/B_n)$

$$\text{Total } P(A) = \sum_{i=1}^{n} P(B_i) \cdot P(A/B_i)$$

$$\therefore P\left(B_1/A\right) = \frac{P(B_1) \cdot P(A/B_1)}{\sum_{i=1}^{n} P(B_i) \cdot P(A/B_i)}$$

071 shrawn + 071 chaitra.

8) Define multiplication law of probablity for dependent & independent events with suitable examples. The independent probablities that the three sections of a costing department will encounter a computer error 0.2, 0.3 & 0.1 per week respectively. What is the probablity that there would be:

i) At least one computer error per week?

ii) One and only one computer error per week?

~~Soln~~ ⇒ If 'A' & 'B' are two dependent event.

then, $P(A \text{ and } B) = P(A \cap B) = P(A) \cdot P(B/A)$, $P(A)>0$

$= P(B) \cdot P(A/B)$, $P(B)>0$

& If 'A' & 'B' are two independent events

then $P(A/B) = P(A)$ & $P(B/A) = P(B)$

so, $P(A \cap B) = P(A) \cdot P(B)$

$= P(B) \cdot P(A)$

For example.

let a die be thrown & let us define the event

$A = \{1,2,3\}$, $B = \{1,3,5\}$, $C = \{3,6\}$, $D = \{2,4,6\}$

Then,

$$P(A) = \frac{O(E)}{O(S)} = \frac{1}{2} = P(B) = P(D);\ P(C) = 1/3$$

$$\text{Now, } P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{2/6}{3/6} = \frac{2}{3}$$

Thus $P(A/B) \neq P(A)$

so, A & B are dependent

Also, $P(C/D) = \frac{P(C\cap D)}{P(D)} = \frac{1/6}{3/6} = 1/3$

$\therefore P(C/D) = P(C)$

so, C & D are independent

Numerical part

Soln Given, Independent events

$P(A) = 0.2, P(B) = 0.3, P(C) = 0.1$

i) At least one computer error per week

$P(A\cup B\cup C) = P(A) + P(B) + P(C) - P(A\cap B) - P(B\cap C) - P(A\cap C) + P(A\cap B\cap C)$

$= P(A) + P(B) + P(C) - P(A).P(B) - P(B).P(C) - P(A).P(C) + P(A).P(B).P(C)$

$= 0.2 + 0.3 + 0.1 - 0.2\times0.3 - 0.3\times0.1 - 0.2\times0.1 + 0.2\times0.3\times0.1$

$= 0.496$

ii) One and only one computer error per week

$P(A\cap \bar{B}\cap \bar{C}) + P(\bar{A}\cap B\cap \bar{C}) + P(\bar{A}\cap \bar{B}\cap C)$

$= P(A).P(\bar{B}).P(\bar{C}) + P(\bar{A}).P(B).P(\bar{C}) + P(\bar{A}).P(\bar{B}).P(C)$

$= 0.2\times0.7\times0.9 + 0.8\times0.3\times0.9 + 0.8\times0.7\times0.1$

$= 0.398$

chaitra +072 chaitra

Q) Define independent and mutually exclusive events with an example. An assembly plant receives its voltage regulators from these three different suppliers, 60% from supplier

A, 30% from supplier B and 10% from supplier C. It is also known that 95% of voltage regulators from A, 80% of these from B and 65% of these from C perform according to specifications. What is the probablity that:-

i) Anyone voltage regulator received by the plant will perform according to specifications

ii) A voltage regulator that perform according to specification came from B & C.

⇒ let us take a two events 'A' & 'B'

If $P(A \text{ and } B) = P(A \cap B) = P(A/B) \cdot P(B) = P(A) \cdot P(B)$

so, the events A & B are independent.

For eg.

let a die be thrown & let us define the event

$A = \{3, 6\}$, $B = \{2, 4, 6\}$

Then, $P(A/B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{1/6}{3/6} = 1/3$

$P(A) = 2/6 = 1/3$

$\therefore P(A/B) = P(A)$

Again,

If $P(A \text{ and } B) = P(A \cap B) = \phi$

then it is said to be mutually exclusive event

For eg.

let a coin is tossed

then $P(H) = 1/2$, $P(T) = 1/2$

$P(H \cup T) = P(H) + P(T) - P(H \cap T) = 1/2 + 1/2 - 0 = 1$

Numerical part

Sol^n Given, $P(A) = 60\% = 0.60$

$P(B) = 30\% = 0.30$

$P(C) = 10\% = 0.10$

if x is the product that meet the specification,

then, $P(x/A) = 95\% = 0.95$

$P(x/B) = 80\% = 0.80$

$P(x/C) = 65\% = 0.65$

i) $P(x) = ?$

ii) $P(B/x) = ?$ & $P(C/x) = ?$

Now,

we have.

i) $P(x) = P(A \cap x) + P(B \cap x) + P(C \cap x)$

$= P(A) \cdot P(x/A) + P(B) \cdot P(x/B) + P(C) \cdot P(x/C)$

$= 0.60 \times 0.95 + 0.30 \times 0.80 + 0.10 \times 0.65$

$= \underline{\underline{0.875}}$

ii) a) $P(B/x) = \dfrac{P(B \cap x)}{P(x)}$

$= \dfrac{P(B) \cdot P(x/B)}{P(x)}$

$= \dfrac{0.30 \times 0.80}{0.875}$

$= \underline{\underline{0.2743}}$

ii) b) $P(C/x) = \dfrac{P(C \cap x)}{P(x)}$

$$= \frac{P(C) \cdot P(x/c)}{P(x)}$$

$$= \frac{0.10 \times 0.65}{0.875}$$

$\therefore P(c/x) = 0.0743.$

Hence, i) probablity anyone voltage regulator received by the plant will perform according to specifications is 0.875.

ii) the probablity that a voltage regulator that perform according to specification came from B & C are 0.2743 & 0.0743 respectively.

chapter:- 2 (Discrete Probablity Distributions)

2.1) Descrete random variable
A discrete random variable is an random variable whose possible values either constitute a finite set or else can be listed in an infinite sequence in where there is a first element, a second element, & so on.

2.2) Binomial Probablity Distributions.
* Condition for Binomial Distribution.
- Number of trail are finite.
- Each trail has two outcomes i.e head or tail, success or failure, pass or fail, etc.
- The probablity of success in each trail remains same.
  eg. probablity of getting head = 1/2, probablity of getting tail = 1/2.
- sum of probablity of success & failure is unity.
- The outcome of different trail are independent.

* Definition
Let X be the random variable with parameter n ; p is said to follow Binomial distribution if it assume non-negative value & its probablity mass function is given by,

$$P(X=x) = P(x) = \begin{cases} {}^nC_x\, p^x q^{n-x} & x = 0, 1, 2, \dots n \\ 0 & \text{otherwise} \end{cases}$$

where,
n = no. of trail (or degree of distribution).

Q) write two conditions that a function is a probablity mass function. ch. 2

p = probablity of success
q = Probablity of failure
& $^nC_x = \frac{n!}{(n-x)!\,x!}$

The mean of the distribution $= \mu = E(X) = np$
The variance of the distribution $= \sigma^2 = V(X) = npq$

Note:- $n \geq 5\%$ of N.

Q) A unbaised coins is flipped 20 times. Find the probablity of getting.
i) Exactly a 7 head
ii) At least 4 head
iii) At most 18 head

soln let x = number of heads
Given, number of trials $(n) = 20$
Probablity of success $(p) = 1/2$
Probablity of failure $(q) = 1-p = 1-1/2 = 1/2$

i) $x = 7$

$\therefore p(X=x) = p(x) = {}^nC_x \, p^x q^{n-x}$

$= {}^{20}C_7 \left(1/2\right)^7 \left(1/2\right)^{20-7}$

$= \underline{\underline{0.0739}}$

ch. 2

ii) $x \geq 4$

$$P(X \geq x) = 1 - P(x < 4)$$
$$= 1 - [P(0) + P(1) + P(2) + P(3)]$$
$$= 1 - \left[(1/2)^{20} + {}^{20}C_1 (1/2)(1/2)^{19} + {}^{20}C_2 (1/2)^2 (1/2)^{18} + {}^{20}C_3 (1/2)^3 (1/2)^{17} \right]$$
$$= 0.9987$$

iii) $x \leq 18$

$$P(X \leq 18) = 1 - P(X > 18)$$
$$= 1 - [P(19) + P(20)]$$
$$= 1 - \left[{}^{20}C_{19} (1/2)^{19} (1/2)^1 + {}^{20}C_{20} (1/2)^{20} \right]$$
$$= 0.99998$$

071 chaitra
072 Ashadh (5) A quality control engineer inspects a random sample of 4 batteries from each lot of 24 car batteries that is ready to shipment. If such a lot contains six (x) batteries with slight defects. What are the probabilities that the inspector's sample will contain,

distribution

paxo ho garda

lost

bala

hypergeometric

i) None of the batteries with defect?

ii) At least two of the batteries with defects?

iii) At most three of the batteries with defects?

soln let x = random variable denoting defects

Given, number of trials $(n) = 4$

ch. 2

probablity of defects $(p) = 6/24 = 1/4$
probablity of undefects $(q) = 1-p = 1-1/4 = 3/4$

i) $x = 0$

$\therefore P(X=x) = P(X=0) = {}^nC_x \, p^x \, q^{n-x}$

$= {}^4C_0 \, (1/4)^0 (3/4)^{4-0}$

$= 1 \times 1 \times (3/4)^4$

$= 0.316$

ii) $x \geq 2$

$\therefore P(X \geq 2) = 1 - (P(0) + P(1))$

$= 1 - \left[0.316 + {}^4C_1 \, (1/4)^1 (3/4)^{4-1}\right]$

$= 1 - [0.316 + 0.421875]$

$= 0.262125$

iii) $x \leq 3$

$\therefore P(X \leq 3) = P(0) + P(1) + P(2) + P(3)$

$= P(0) + P(1) + {}^4C_2 \, (1/4)^2 (3/4)^2 + {}^4C_3 \, (1/4)^3 (3/4)^1$

$= 0.316 + 0.421875 + 0.2109375 + 0.046875$

$= 0.9956875$

061 Ashwin
067 mangsir

Q) If 20% of the bolt produced by a machine are defective, find the probablity that out of 4 bolts ~~ea~~ choosen random
i) Exactly one bolt will be defective.
ii) No defective bolt
iii) less than 2 bolts will be defective.

Sol^n let x = random variable denoting no. of defects
Given, no. of trials $(n) = 4$
probablity of success $(P) = 20/100 = 1/5$
probablity of failure $(q) = 1-p = 1-1/5 = 4/5$

i) $X=1$

$$\therefore P(X=x) = P(X=1) = {}^nC_x p^x q^{n-x}$$
$$= 4C_1 (1/5)^1 (4/5)^{4-1}$$
$$= 0.4096.$$

ii) $X=0$

$$\therefore P(X=x) = P(x=0) = {}^nC_x p^x q^{n-x}$$
$$= 4C_0 (1/5)^0 (4/5)^{4-0}$$
$$= 0.4096.$$

iii) $\cancel{X \geq 2}$ $X \leq 2$

$$\therefore \cancel{P(X \geq 2) = 1 - (P} \; P(X \leq 2) = P(0) + P(1)$$
$$= 0.4096 + 0.4096$$
$$= 0.8192$$

$$p = {}^nC_x\, p^x q^{n-x}$$

ch. 2

Q) At a particular university it has been found that 20% of the student withdraw without completing BE course. Assume that 18 students have registered for the course the semester.

a) What is the probablity that none will withdraw?

b) What is the probablity that at least one will withdraw?

c) What is the probablity that at most 2 will withdraw?

solⁿ let x = random variable denoting withdrawn students.

Given, no. of trials / students $(n) = 18$

probablity of success $(p) = 20/100 = 1/5$

probablity of failure $(q) = 1-p = 1-1/5 = 4/5$

i) $X=0$

$\therefore p(X=0) = p(x=0) = p(0) = {}^nC_x\, p^x q^{n-x}$

$= {}^{18}C_0 (1/5)^0 (4/5)^{18-0}$

$= 0.018$

ii) $X \geq 1$

$\therefore p(X\geq 1) = \cancel{p(x=1) = p(1) = {}^nC_x\, p^x q^{n-x}} = 1 - p(0)$

$\cancel{= {}^{18}C_1 (1/5)^1 (4/5)^{18-1}} = 1 - 0.018$

$= 0.982.$

iii) $X \leq 2$

$\therefore p(X\leq 2) = p(0) + p(1) + p(2)$

$= p(0) + {}^{18}C_1 (1/5)^1 (4/5)^{18-1} + {}^{18}C_2 (1/5)^2 (4/5)^{18-2}$

$= 0.2713$

Note: $b(x, n, p) = \Sigma\, {}^nC_x p^x q^{n-x}$

ch. 2

8) Multiple choice test consist of 10 question & 4 answer to each question. If each question is answer by shifting 4 toys labeled 1, 2, 3 & 4 drawing one & making the alternate whose number is drawn. Find the probability of
a) getting of 3 correct answer?
b) At least 1 correct answer?
c) At most 3 of these answered correctly?
d) At least 7 correct answer?

Sol^n: let x = random variable denote the correct answer.
Given no. of trials (questions) $(n) = 10$
probability of success $(p) = 1/4$
probability of failure $(q) = 1-p = 1-1/4 = 3/4$

i) $X = 3$

$\therefore P(X=x) = P(x=3) = P(3) = {}^nC_x p^x q^{n-x}$
$= {}^{10}C_3 (1/4)^3 (3/4)^{10-3}$
$= 0.25028$

ii) $X \geq 1$

$\therefore P(X \geq 1) = 1 - P(0)$
$= 1 - {}^nC_x p^x q^{n-x}$
$= 1 - {}^{10}C_0 (1/4)^0 (3/4)^{10-0}$
$= 0.9436$

iv) $X \geq 7$

$\therefore P(X \geq 7) = P(7) + P(8) + P(9) + P(10)$

ch. 2

$= {}^{10}C_7 (1/4)^7 (3/4)^3 + {}^{10}C_8 (1/4)^8 (3/4)^2 + {}^{10}C_9 (1/4)^9 (3/4)^1 + {}^{10}C_{10} (1/4)^{10} (3/4)^0$

$= 0.0035057$

iii) ~~$X \geq 3$~~ $X \leq 3$

$\therefore P(X \leq 3) = P(0) + P(1) + P(2) + P(3)$

$= {}^{10}C_0 (1/4)^0 (3/4)^{10} + {}^{10}C_1 (1/4)^1 (3/4)^9 + {}^{10}C_2 (1/4)^2 (3/4)^8 + {}^{10}C_3 (1/4)^3 (3/4)^7$

$= 0.775875$

2063 Kartik

8) If 90% of all students taking a begining computer course fail to get their first program to run of first submission What is the probability that among 15 randomly choosen such student?

a) At least 12 will fail on first submission.

b) Between 10 & 13 inclusive will fail on first submission.

c) At most 2 get their program will to run properly on first submission.

soln let x = random variable denoting fail students

Given, no. of trial $(n) = 15$

probablity of success $(p) = 90/100 = 9/10$

probablity of failure $(q) = 1 - p = 1 - 9/10 = 1/10$

a) At least 12 will fail $(X \geq 12)$

$P(X \geq 12) = P(12) + P(13) + P(14) + P(15)$

$= {}^{15}C_{12} (9/10)^{12} (1/10)^3 + {}^{15}C_{13} (9/10)^{13} (1/10)^2 + {}^{15}C_{14} (9/10)^{14} (1/10)^1 +$

ch. 2

$^{15}C_{15} (9/10)^{15} (1/10)^0$
$= 0.999999.$

b) Between 10 & 13 inclusive $(10 \le X \le 13)$

$p(10 \le X \le 13) = p(10) + p(11) + p(12) + p(13)$

$= {}^{15}C_{10} (9/10)^{10} (1/10)^5 + {}^{15}C_{11} (9/10)^{11} (1/10)^4 + {}^{15}C_{12} (9/10)^{12} (1/10)^3 + {}^{15}C_{13} (9/10)^{13} (1/10)^2$

$= 0.4497$

c) At most 2 $(X \le 2)$

$P(X \le 2) = P(0) + P(1) + P(2)$

$= {}^{15}C_0 (9/10)^0 (1/10)^{15} + {}^{15}C_1 (9/10)^1 (1/10)^{14} + {}^{15}C_2 (9/10)^2 (1/10)^{13}$

$= 8.641 \times 10^{-12}$

2.3) Negative binomial distribution

The negative binomial distribution depends upon the binomial distribution but the last trail must be success in negative binomial distribution. In negative binomial distribution the experiments is repeated until success occures.

Consider a random experiment 'X' with probablity mass function (P.m.F)

$f(x, r, p)$ where,

x = no. of trails

8) Define negative binomial distribution with examples.

ch. 2

r = number of success

p = probablity of success.

The negative binomial distribution is given as,

$$^{x+r-1}C_{r-1}\, p^{r-1}\, q^{x}$$

Therefore by compound probablity theorem $f(x,r,p)$ is given by the product of then two probablities.

i.e. $f(x,r,p) = {}^{x+r-1}C_{r-1} \cdot p^{r-1} \cdot q^{x} \cdot p$

$$f(x,r,p) = {}^{x+r-1}C_{r-1}\, p^{r} \cdot q^{x}$$

* Definition

Consider a sequence of independent repetation of random experiments with constant probablity of success "p".

let a random variable x denote the total no. of failures in the experiment before the r^{th} success so that $x+r$ is the total number of trails. Then the probablity density function is given by,

$$p(X=x) = p(x) = nb(x; n, p)$$

$$= {}^{x+r-1}C_{r-1}\, p^{r} \cdot q^{x}$$

where $x = 0, 1, 2, 3, \ldots$

OR, A descrete random variable 'X' with parameter r & p is said to be in negative binomial distribution if its probablity mass function (p.m.f) is given by

$$p(x) = p(X=x) = {}^{x+r-1}C_{r-1}\, p^{r}\, q^{x}$$

where $x = 0, 1, 2, 3, \ldots$

ch.2

* Mean & variance

Mean $E(x) = \frac{r(1-p)}{p} = \frac{rq}{p}$ $(\because 1-p=q)$

variance $V(x) = \frac{r(1-p)}{p^2} = \frac{rq}{p^2}$

* characteristics

- Total probablity of negative binomial distribution is unity.

$$\sum_{x=0}^{\infty} P(X=x) = p^r \sum_{x=0}^{\infty} ({}^{-r}C_x)(-q)^x = 1$$

Therefore $P(X=x) = {}^{-r}C_x\, p^r (-q)^x$ is p.m.f

- If we take $r=1$ in ${}^{x+r-1}C_{r-1}\, p^r q^x$ then we have

$$P(X) = p(X=x) = q^x \cdot p \text{ where } x = 0, 1, 2, \ldots$$

This is known as probablity function of geometric distribution.

Q) If a boy is throwing stone at a target what is the probablity that his 10^{th} throw is his 5^{th} hit, if the probablity of hitting the target at any trial is ½.

soln let x = random variable not hitting a target

Given, probablity of hitting a target $(p) = \frac{1}{2}$, $q = 1 - \frac{1}{2} = \frac{1}{2}$

Total number of throws $= x + r = 10$

Number of success $(r) = 5$

Now, by using negative binomial distribution,

$$P(X=x) = P(5) = {}^{x+r-1}C_{r-1}\, p^r \cdot q^x$$

$$= {}^{10-1}C_{5-1}\, p^5 q^5$$

$$= {}^{9}C_4 (1/2)^5 (1/2)^5$$

$$= 0.123.$$

PU 2001

2) An item is produced in large number. The machine is known to produce 5% defective. A quality control ~~isp~~ inspector is examing that the item by taking them of random. What is the probability that at least 4 items are to be examined in order to get 2 defectives?

Soln let x = random variable denoting defective items.

propablity of producing defective item $= 5/100 = 1/20 = p$

Total no. of sample $= x + r = 4$

Number of success $= r = 2$ $\therefore$ ~~x=2~~

Now,

If 2 defective items are to be obtained from 4 items then,

$$p(X \geq 2) = 1 - (p(0) + p(1))$$

$$= 1 - {}^{x+r-1}C_{r-1}\, p^r q^x - {}^{x+r-1}C_{r-1}\, p^r q^x$$

$$= 1 - {}^{3}C_1\, p^2 q^0 - {}^{3}C_1\, p^2 q^1$$

$$= 1 - {}^{3}C_1 (1/20)^2 (19/20)^0 - {}^{3}C_1 (1/20)^2 (19/20)^1$$

$$= 0.99275.$$

ch.2

8) Different between Binomial Distribution & Negative Binomial Distribution.

Sol^n

Binomial Distribution	Negative Binomial Distribution.
1) Binomial Distribution is based on Bernoulli's trail such as pass or fail, male or female, etc.	1) Negative Binomial Distribution is based on the binomial distribution but the last trail must be success in negative binomial distribution.
2) The no. of trails in binomial distribution is finite and fixed.	2) The no. of trails are repeated until a desirable number of success occurs.
3) In binomial, the no. of trails is fixed and the no. of successes to ~~ee~~ occurs is random variable.	3) But in negative binomial distribution the number of successes is fixed and the number of trails is a random variable.
4) For example, Probablity of 5 even ~~e~~ numbers when dice is thrown 30 times.	4) For example, 30 throw of a dice will sufficient to generate even number.

Ch. 2

Q) Difference between the Binomial & Hypergeometric distribution

⇒

Binomial Distribution	Hypergeometric Distribution
1) In binomial distribution there is independent of trial so sampling with replacement is done.	1) But in Hypergeometric distribution sampling without replacement is done.
2) It is applicable, where the probablity of success is same for all trials.	2) If the probablity of success varies from trial to trial then hypergeometric distribution is suitable.
3) Binomial distribution is the approximate probablity model for sampling from a finite dichotomus population.	3) The hypergeometric distribution is the exact probablity model for the number of success is same sample.

2.4) Poison distribution

Poisson distribution is limiting case of binomial distribution under the following conditions.

- No. of trail 'n' is indefinitely large i.e $n \to \infty$
- The probablity of success in each trail is constant

Generally Poisson distribution have $n \geq 50$ & $p < 0.1$.

Ch.2

and is indefinitely small i.e $p \to 0$

- $np = m$ (say) is finite ; mean $= m =$ variance, $E(X) = m = V(X)$

* Definition

A random variable 'X' is said to follow a poisson distribution if it has small probablity ($p \to 0$) with large sample ($n \geq 100$) and its probablity mass function is given by,

$$p(X=m) = p(X=x) = \frac{e^{-m} m^{x}}{x!} \quad ; x = 0, 1, 2, 3, \dots \infty$$

$$= 0 \quad \text{otherwise.}$$

where, m is known as the parameter of the distribution.

* Some examples

- No. of telephone calls arriving at telephone switch board in unit time.
- No. of accident only Air lbus.
- No. of sucidel death of certain diseases in certain time.

* Poisson approximation to the Binomial distribution

OR, Limiting case as binomial distribution.

The number of items more than 20 and probablity of the item is less than 10%, It is better to use poisson distribution instead of binomial distribution, however, they can be solved by binomial distribution.

ch.2

The poisson distribution is the limiting case of Binomial distribution under following conditions.

- If the number of trail is indefinitely large ie $n \to \infty$
- Probablity of success for each trail is infinite small ie $p = 0$.

In mathematical calculation of such condition we use poisson distribution with parameter, $m = np$.

* Properties of Poisson Distribution.

i) Total probablity is unity

$$\sum_{x=0}^{\infty} P(X=x) = \sum_{x=0}^{\infty} \frac{e^{-m} m^x}{x!} = 1$$

ii) It is the only distribution known so far of which the mean & variance are equal. so it is uniparametric distribution.

iii) 'm' is called the parameter of poisson distribution, which is average no. of success per unit.

iv) If value of 'λ' or 'm' is known, poisson distribution is known.

v) The probablity that a success will occurs in any interval is same or all intervals of equal size & is proportional to size of interval.

*Theorem

The poisson distribution is the limiting case of binomial distribution with $n \to \infty$ & $p \to 0$ when $np = m > 0$

Proof:-

We have,

$$b(x; n, p) = {}^{n}C_{x}\, p^{x} q^{n-x} \quad ; q = 1-p,\ x = 0, 1, 2, 3, \ldots$$

$$= \frac{n!}{(n-x)!\,x!} \left(\frac{m}{n}\right)^{x} \left(1-\frac{m}{n}\right)^{n-x} \qquad \left[\because np = m\right]$$

$$= \frac{n(n-1)(n-2)\ldots(n-x-1)(n-x)(n-x+1)\ldots 2.1}{x!\,(n-x)!\; n^{x}} . m^{x} \left(1-\frac{m}{n}\right)^{-x} \left(1-\frac{m}{n}\right)^{n}$$

$$= \frac{n(n-1)(n-2)\ldots(n-x-1)(n-x)!}{x!\,(n-x)!\; n^{x}} . m^{x} . \left(1-\frac{m}{n}\right)^{-x} \left[\left(1-\frac{m}{n}\right)^{-n/m}\right]^{-m}$$

$$= \frac{1(1-1/n)(1-2/n)\ldots\left(1-\frac{x-1}{n}\right)}{x!} . m^{x} \left(1-\frac{m}{n}\right)^{-x} \left[\left(1-\frac{m}{n}\right)^{-n/m}\right]^{-m}$$

Now, as $n \to \infty$

$$\lim_{n\to\infty} \left[(1-1/n)(1-2/n)\ldots\left(1-\frac{x-1}{n}\right)\left(1-\frac{m}{n}\right)^{-x}\right] \to 1$$

$$\& \lim_{n\to\infty} \left[\left(1-\frac{m}{n}\right)^{-n/m}\right]^{-m} \to e^{-m}$$

$$\therefore b(x; n, p) = \frac{m^{x} . e^{-m}}{x!}$$

proved

ch.2

Q) An office switch board receive telephone calls at the rate of 3 calls per min on average. What is the probablity of receiving,

a) No call in 1 min. interval.

b) At least 3 calls in one minute.

c) At most 3 calls in a 5 min. interval.

soln let X = random variable denoting all recieves in 1 minute.

poisson distribution parameter (m) = 3

a) probablity of no calls in one minute interval

$$p(X=0) = \frac{e^{-3}3^0}{0!} = e^{-3} = 0.049787$$

b) At least 3 calls in one minute

$$p(X \geq 3) = 1 - [p(0) + p(1) + p(2)]$$

$$= 1 - \left[\frac{e^{-3}3^0}{0!} + \frac{e^{-3}3^1}{1!} + \frac{e^{-3}3^2}{2!}\right]$$

$$= 0.5768.$$

may not be correct

X c) At most ~~3 calls~~ in 5 min interval ~~ie 0.6 in 1 min interval~~

~~first At most 3 calls in 1 min interval~~ m = 3 in 1 min

i.e m = 15 in 5 min

$$p(X \leq 3) = p(3) + p(0) + p(1) + p(2)$$

$$= \frac{e^{-15}15^3}{3!} + \cancel{= 0.224} \frac{e^{-15}15^0}{0!} + \frac{e^{-15}15^1}{1!} + \frac{e^{-15}15^2}{2!} = \cancel{0.6472} = 0.00021$$

∴ ~~probablity in 5 min interval = 0.224 × 5 0.64725 × 5 =~~

~~= 1.12 = 3.23.~~

5) An office switch board receive a telephone called at rate of 3 calls per minute on average. Find the probablity of receiving
i) No call in one minute
ii) At most 2 calls in five minutes interval.

soln let X = random variable denoting no. of call received in 1 min.
poisson distribution parameter $(m) = 3$

a) probablity of no calls in one minute

$$p(X=0) = \frac{e^{-m} m^{x}}{x!} = \frac{e^{-3} 3^{0}}{0!} = e^{-3} = 0.049787.$$

b) At most 2 calls in five minutes interval
first, at most 2 calls in 1 minute interval

$$p(X \leq 2) = p(0) + p(1) + p(2)$$

$$= \frac{e^{-3} 3^{0}}{0!} + \frac{e^{-3} 3^{1}}{1!} + \frac{e^{-3} 3^{2}}{2!}$$

$$= 0.42319$$

ch.2

2.5) Hyper geometric distributions

Sampling without replacement is associated with hypergeometric distribution.

let a lot of N contain M defective & N-M non-defective. If a sample of size n is drawn. Then the hypergeometric distribution for defective is given by,

$$P(X=x) = h(x; n, M, N) = \frac{{}^{M}C_{x} \; {}^{N-M}C_{n-x}}{{}^{N}C_{n}}$$

where $x = 0, 1, 2, 3, \ldots n$

* Mean & Variance of Hypergeometric Distribution

Let hypergeometric random variable X having probability mass function (pmf) $h(x; n, M, N)$ then,

$$\text{Mean} = E(X) = n\frac{M}{N} = n.p$$

$$\text{Variance} = V(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{M}{N}\left(1-\frac{M}{N}\right) = \left(\frac{N-n}{N-1}\right) np(1-p)$$

where $p = \frac{M}{N}$

Note:- $n < \frac{N}{10}$ or $\frac{n}{N} < 0.1$.

ass chaitra

Q) A cartoon of 24 hand grenades contains 4, that are defective If 3 hand grenades are randomly selected from this cartoon, what is the probablity that exactly 2 of them

Hypergeometric distributions $P(X=x) = \dfrac{{}^{M}C_x \; {}^{N-M}C_{n-x}}{{}^{N}C_n}$

M = defective, N = total

n = no. of selection.

are defective?

Sol^n let X = random variable associates with defective land grenades.

Given, Total no. of grenades, $N = 24$

no. of defective grenades, $M = 4$

no. of non-defective grenades $= N-M = 24-4 = 20$

No. of selection $(n) = 3$

Now,

$$P(X=2) = \frac{{}^{M}C_x \; {}^{N-M}C_{n-x}}{{}^{N}C_n}$$

$$= \frac{{}^{4}C_2 \; {}^{20}C_1}{{}^{24}C_3} = 0.059288$$

Q) From a lot containing 25 items, 5 of which are defective, 4 are choosen at random. Obtain the probablity distribution of a number of defective items drawn. Also find the probablity of 3 or 4 defective items.

Sol^n let X = random variable denoting defective items

Given, Total number $(N) = 25$

No. of defective $(M) = 5$

No. of non-defective $(N-M) = 25-5 = 20$

No. of selection $(n) = 4$

$mean = E(X) = np = n \cdot \frac{M}{N}$

Now, we know by hypergeometric probablity distribution is

$$p(X=x) = \frac{{}^{M}C_{x}\ {}^{N-M}C_{n-x}}{{}^{N}C_{n}}$$

So, probablity of 3 or 4 defective

$$= p(X=3 \text{ or } X=4) = p(X=3 \cup X=4)$$

$$= p(X=3) + p(X=4)$$

$$= \frac{{}^{5}C_{3}\ {}^{20}C_{1}}{{}^{25}C_{4}} + \frac{{}^{5}C_{4}\ {}^{20}C_{0}}{{}^{25}C_{4}}$$

$$= \underline{\underline{0.0162}}$$

Q) A shipment of 20 tape recorder contains 5 that are defective. If 10 of them are randomly choosen for Impection, What is the probablity that 2 of 10 will be defective also find the mean & variance of distribution.

Soln. let X = random variable denoting defective tape

Given, Total no. of tape (N) = 20

no. of defective (M) = 5

no. of non-defective (N−M) = 20−5 = 15

no. of selection (n) = 10

Now,

Using hypergeometric distribution

for X = 2, $p(X=2) = \frac{{}^{M}C_{x}\ {}^{N-M}C_{n-x}}{{}^{N}C_{n}}$

variance $= V(x) = \left(\frac{N-n}{N-1}\right) np(1-p) = \left(\frac{N-n}{N-1}\right) n \cdot \frac{M}{N}\left(1-\frac{M}{N}\right)$

N = total no. M = no. of defective.

$$= \frac{{}^{5}C_2 \; {}^{15}C_8}{{}^{20}C_{10}} = \underline{\underline{0.348297.}}$$

Also,

$$\text{Mean} = E(X) = \cancel{np(x)} = np = \frac{n \cdot M}{N} = \frac{10 \times 5}{20} = \underline{\underline{2.5}}$$

$$\text{variance} = V(x) = \left(\frac{N-n}{N-1}\right) np(1-p)$$

$$= \left(\frac{N-n}{N-1}\right) n \cdot \frac{M}{N}\left(1-\frac{M}{N}\right)$$

$$= \left(\frac{20-10}{20-1}\right) 10 \times \frac{5}{20}\left(1-\frac{5}{20}\right)$$

$$= \cancel{0.9862}\ \underline{\underline{0.9868}}$$

5) A quality control Engineer inspect a random sample of 4 batteries from each lot of 24 car batteries ready to be shipped. If such a lot contains 6 batteries with sight defects, what are the probablity that the inspected sample will contains

a) None of the batteries with defects?

b) only of the batteries with defects?

c) At least 2 of batteries with defects?

d) At most three of the batteries with defects?

Soln let x = random variable denoting defect batteries

no. of batteries (N) = 24

ch.2

no. of slight defect batteries $(M) = 6$
no. of non-defect batteries $(N-M) = 24-6 = 18$
no. of selection $(n) = 4$

Now,
using hypergeometric distribution,

$$P(X=x) = \frac{{}^{M}C_{x} \; {}^{N-M}C_{n-x}}{{}^{N}C_{n}}$$

a) $P(X=0)$

$$= \frac{{}^{6}C_{0} \; {}^{18}C_{4}}{{}^{24}C_{4}}$$

$$= 0.28797$$

b) $P(X=1)$

$$= \frac{{}^{6}C_{1} \; {}^{18}C_{3}}{{}^{24}C_{4}}$$

$$= 0.4607566$$

c) $P(X \geq 2)$

$$= P(2) + P(3) + P(4)$$

$$= \frac{{}^{6}C_{2} \; {}^{18}C_{2}}{{}^{24}C_{4}} + \frac{{}^{6}C_{3} \; {}^{18}C_{1}}{{}^{24}C_{4}} + \frac{{}^{6}C_{4} \; {}^{18}C_{0}}{{}^{24}C_{4}}$$

$$= \cancel{0.1304} \; 0.25127$$

d) $P(X \leq 3)$

$$= 1 - P(4) = 1 - \frac{{}^{6}C_{4} \; {}^{18}C_{0}}{{}^{24}C_{4}} = 0.998588$$

ch.2

069 Bhadra

Q) If 16 of 18 new building in a city voilate the building code what is the probability that the building inspector who randomly selects 4 of the new building for inspection will catch.
a) None of the building that violate the building code.
b) One of the building that violate the building code

Soln let X = random variable denoting building with code.
Given, No. of building (N) = 18
Building with code (M) = 16
Building without code (N - M) = 18 - 16 = 12
no. of selection (n) = 4

Now,
Using hypergeometric distribution,

$$p(X=x) = \frac{{}^{M}C_{x}\ {}^{N-M}C_{n-x}}{{}^{N}C_{n}}$$

a) $p(X=0)$

$$= \frac{{}^{16}C_{0}\ {}^{12}C_{4}}{{}^{18}C_{4}}$$

$= 0.16176$

b) $p(X=1)$

$$= \frac{{}^{6}C_{1}\ {}^{12}C_{3}}{{}^{18}C_{4}}$$

$= 0.4313$

ch. 2

Q) A taxi cab has 12 Ambassador and 8 Fiats. If 5 of these taxi cabs are in the shop for repairs and Ambassador is as likely to be in for repairs as a fiat, what is the probablity that,

i) 3 of them are Ambassador and 2 are Fiats?

ii) 3 of them are Fiats & 2 are Ambassador?

iii) At least 3 of them are Ambassador?

iv) All 5 of them are of the same make?

Soln let X = random variable denoting ambassador

Total no. of taxi = N = 12+8 = 20

No. of ambassador (M) = 12

No. of fiats (N−M) = 20−12=8

no. of selection (n) = 5

Now,

using hypergeometric distribution,

$$p(X=x) = \frac{{}^{M}C_{x}\ {}^{N-M}C_{n-x}}{{}^{N}C_{n}}$$

i) Probablity that 3 of them are ambassador & 2 are fiats

$p(X=3)$

$$= \frac{{}^{12}C_3\ {}^{8}C_2}{{}^{20}C_5}$$

$$= 0.3973$$

ch. 2

ii) probablity that 3 of them are fiats & 2 are ambassador

$p(X=2)$

$$= \frac{{}^{12}C_2 \, {}^{8}C_3}{{}^{20}C_5}$$

$$= 0.23839$$

iii) Probablity that at least 3 of them are ambassador

$$p(X \geq 3) = p(3) + p(4) + p(5)$$

$$= \frac{{}^{12}C_3 \, {}^{8}C_2}{{}^{20}C_5} + \frac{{}^{12}C_4 \, {}^{8}C_1}{{}^{20}C_5} + \frac{{}^{12}C_5 \, {}^{8}C_0}{{}^{20}C_5}$$

$$= 0.7038.$$

iv) Probablity that all 5 of them are of same make i.e either all are ambassador or all are fiat.

$p(X=5 \text{ or } Y=5)$

$= p(X=5) + p(Y=5)$

$$= \frac{{}^{12}C_5 \, {}^{8}C_0}{{}^{20}C_5} \cancel{\frac{{}^{12}C_5 \, {}^{8}C_0}{{}^{20}C_5}} + \frac{{}^{12}C_0 \, {}^{8}C_5}{{}^{20}C_5}$$

$$= 0.05469556$$

Questions from question bank.

Q) A heavy machinery manufacture has 3840 large generators in the field that are under warrenty. If the probablity is 1/1200 that any one will fail during the given

ch.2

year, find the probablity
i) That exactly 3 generators will fail during the given year?
ii) That between 2 & 6 are fail during the given year?

Soln X = random variable denoting failure generators
poisson distribution parameter $(m) = np = 3840 \times \frac{1}{1200}$
$= 3.2$

Now,
using poisson distribution
$$p(X=x) = \frac{e^{-m} m^x}{x!}$$

i) 3 generator will fail during the given year,
$$p(X=3) = \frac{e^{-3.2}(3.2)^3}{3!}$$
$= 0.2226159833$

ii) between 2 & 6 are fail during the given year
$$p(2 \le X \le 6) = p(2) + p(3) + p(4) + p(5) + p(6)$$
$$= \frac{e^{-3.2}(3.2)^2}{2!} + \frac{e^{-3.2}(3.2)^3}{3!} + \frac{e^{-3.2}(3.2)^4}{4!} + \frac{e^{-3.2}(3.2)^5}{5!} + \frac{e^{-3.2}(3.2)^6}{6!}$$
$= 0.7841796423$

prochaitra
Q) The number of accident in a year attributes to taxi drivers in a city follows Poisson distribution with mean 3

49.7870

ch.2

Out of 1000 taxi driver, find the approximately the number of driver with.

i) No accidents in a year

ii) More than 3 accident in a year.

sol^n let X = random variable denoting accident in a year

poisson distribution parameter $(m) = 3$

no. of taxi drivers $(n) = 1000$

Now,

using poisson distribution

$$p(X=x) = \frac{e^{-m} m^{x}}{x!}$$

i) No accidents in a year $(X=0)$

$$\therefore p(X=x) = p(X=0) = \frac{e^{-3} 3^{0}}{0!}$$

$$= 0.04979$$

$\therefore$ No. of taxi driver with no accident is

$$1000 \times p(X=0)$$

$$= 1000 \times 0.04979$$

$$= 49.79 \approx 50$$

ii) More than 3 accident in a year $(X \geq 3)$

$$p(X \geq 3) = 1 - [P(0) + P(1) + P(2)] + P(3)$$

$$= 1 - \left[\frac{e^{-3} 3^{0}}{0!} + \cancel{\frac{e^{-3} 3^{1}}{1!}} + \frac{e^{-3} 3^{2}}{2!}\right]$$

$$= 0.57681$$

random भनो भने binomial distribution लगाउने

ch.2

∴ No. of taxi driver with more than 0 accident in a year

$= 1000 \times P(X \geq 3)$

$= 1000 \times 0.57681$

$= 576.81 \approx \underline{\underline{577}}$

chapter:-3 (continuous Probablity Distribution)

* Rectangular or Uniform distribution

Let X be a continuous random variable over an interval (a,b) and its said to be uniform distribution if

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

* characteristics

- Total probablity of unity

i.e. $\int_a^b f(x)\,dx = 1.$

- If f(x) is probablity density function of continuous random variable X then.

$$\frac{d}{dx} F(x) = f(x) \text{ (probablity density function)}$$
$$= \frac{1}{b-a} \neq 0$$

* Mean

$$E(X) = \int_a^b x\,f(x)\,dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b$$

$$= \frac{1}{b-a}\left[\frac{b^2-a^2}{2}\right] = \frac{b+a}{2}$$

so, $E(X) = \frac{b+a}{2}$

* Variance

$$V(X) = E(X^2) - \{E(X)\}^2$$

descrete probablity distribution ⇒ probablity mass Function (pmf)
continuous probablity distribution ⇒ Probablity density Function (pdf).

---

since $E(X^2) = \int_a^b x^2 f(x)\,dx$

$$= \frac{1}{b-a}\int_a^b x^2\,dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b$$

$$= \frac{1}{b-a}\left[\frac{b^3-a^3}{3}\right] = \frac{1}{b-a}\,\frac{(b-a)(b^2+ab+a^2)}{3}$$

$$= \frac{b^2+ab+a^2}{3}$$

$$\therefore V(X) = \frac{b^2+ab+a^2}{3} - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{4(b^2+ab+a^2) - 3(a^2+2ab+b^2)}{12}$$

$$= \frac{4b^2+4ab+4a^2-3a^2-6ab-3b^2}{12}$$

$$= \frac{b^2-2ab+a^2}{12}$$

$$= \frac{(b-a)^2}{12}$$

3.1) Continuous random variable & probablity densities

A random variable X which takes real value in an interval is a continuous random variable & the density function f(x) which satisfy the condition

i) $f(x) \geq 0$

ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

is called the probability density function & the distribution function $F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx \ , \ -\infty < x < \infty$

so that,

$$\frac{d}{dx} F(x) = f(x).$$

chaitra

Q) A random variable X has the following probability density function as:-

$$f(x) = \begin{cases} kx^3(4-x)^2 & , 0 < x < 3 \\ 0 , & \text{otherwise} \end{cases}$$

Find the value of k, using this value of k, find mean & variance of distribution.

Soln Given,

$$f(x) = \begin{cases} kx^3(4-x)^2 & , 0 < x < 3 \\ 0 & , \text{otherwise} \end{cases}$$

Now,

We have for continuous probability distribution

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

or, $\int_{-\infty}^{0} f(x)\,dx + \int_{0}^{3} f(x)\,dx + \int_{3}^{\infty} f(x)\,dx = 1$

or, $0 + \int_{0}^{3} kx^3(4-x)^2\,dx + 0 = 1$

~~or, $k\left[\frac{x^3 \cdot - x^2}{2} + \right.$~~

$$\text{or, } K\left[(4-x)^2 \cdot \frac{x^4}{4} - \int (-2)(4-x) \cdot \frac{x^4}{4}\, dx\right] = 1$$

$$\text{or, } K\left[(4-x^2) \cdot \frac{x^4}{4} - \int \left(\frac{x^5}{2} - 2x^4\right) dx\right] = 1$$

$$\text{or, } K\left[(4-x^2)\frac{x^4}{4} - \left[\frac{x^6}{12} - \frac{2x^5}{5}\right]\right] = 1$$

$$\text{or, } K\left[x^4 - \frac{x^6}{4} - \frac{x^6}{12} + \frac{2x^5}{5}\right]_0^3 = 1$$

$$\text{or, } K\left[81 - \frac{729}{4} - \frac{729}{12} + \frac{486}{5}\right] = 1$$

or,

ch. 3

Q) The probablity density function given by
$f(x) = cx^2, \quad 0<x<3$
$= 0 . \quad$ otherwise

i) Find the value of constant c?
ii) Compute $P(1<x<2)$
iii) Find the distribution function.

Soln Given, $f(x) = cx^2, \quad 0<x<3$
$= 0 \quad .$ otherwise

Now,

We have for continuous probablity distribution

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

or, $\int_{-\infty}^{0} f(x)\,dx + \int_{0}^{3} f(x)\,dx + \int_{3}^{\infty} f(x)\,dx = 1$

or, $0 + \int_{0}^{3} cx^2\,dx + 0 = 1$

or, $\left[\frac{cx^3}{3}\right]_0^3 = 1$

or, $\frac{c \times 3^3}{3} - \frac{c \times 0^3}{0} = 1$

or, $c \times 9 = 1$

or, $c = 1/9$

$\therefore$ i) $\Rightarrow c = 1/9$ //

ii) $\Rightarrow P(a<x<b) = \int_a^b f(x)\,dx$

ch. 3

$$\therefore P(1 < X < 2) = \int_1^2 cx^2\,dx$$

$$= \left[\frac{1}{9}\frac{x^3}{3}\right]_1^2$$

$$= \frac{8}{27} - \frac{1}{27} = \frac{7}{27}$$ //

iii) $F(x) = \int f(x)\,dx$

$$= \int cx^2\,dx$$

$$= c\frac{x^3}{3} + c = \frac{x^3}{27} + c$$ //

070 Ashad

Q) let X denotes the amount of time for which a book on two-hour reserve at a college library is checked out by a randomly selected students, and suppose that X has density function $f(x) = Kx$, $0 \le x \le 2$

0, otherwise

a) find the value of K

b) calculate $P(X \le 1)$

c) calculate $P(0.5 \le X \le 1.5)$

d) calculate $P(1.5 < X)$

soln Given, $f(x) = kx$, $0 \le x \le 2$

0, otherwise

Now,

We have from continuous probablity distribution.

$$\int_{-\infty}^{\infty} f(x) = 1$$

or, $\int_{-\infty}^{0} f(x)dx + \int_{0}^{2} f(x)\,dx + \int_{2}^{\infty} f(x)\,dx = 1$

or, $0 + \int_{0}^{2} kx\,dx + 0 = 1$

or, $\left[\frac{kx^2}{2}\right]_0^2 = 1$

or, $k.\frac{4}{2} = 1$

or, $k = 1/2$

$\therefore$ a) $\Rightarrow$ The value of k is $1/2$.

b) $\Rightarrow$ $P(X \le 1) = \int_{-\infty}^{1} f(x)\,dx$

$= \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx$

$= 0 + \int_{0}^{1} \left(\frac{x}{2}\right) dx$

$= \left[\frac{x^2}{2\times 2}\right]_0^1$

$\therefore P(X \le 1) = \underline{\underline{1/4}}$

ch. 3

c) $P(0.5 \le X \le 1.5) = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{0.5} f(x)\,dx + \int_{0.5}^{1.5} f(x)\,dx + \int_{1.5}^{2} f(x)\,dx + \int_{2}^{\infty} f(x)\,dx$

$=$

ch. 3

3.2) Normal distribution

It is also known as Gaussion distribution. English mathematician De-Maivre (1667 - 1754) discovered normal distribution.

* Definition

Let X be a ~~constant random~~ continuous random variable having mean μ & standard deviation (σ) within the range $(-\infty, \infty)$ is said to be normally distributed value if its probablity density function (p.d.f) is given by,

$$f(X=x) = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$

where μ = mean

σ = standard deviation

Note:-

- Normal distribution are expressed as $X \sim N(\mu, \sigma^2)$ i.e. a random variable X follows normal distribution with mean μ and variance σ^2.

$$\int_{-\infty}^{\infty} e^{-y^2/2}\, dy = \sqrt{2\pi}$$

- Total probablity density is always unity.

$$\int_{-\infty}^{\infty} f(x)\,dx = \frac{\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

put $\frac{x-\mu}{\sigma} = y$

$$\int_{-\infty}^{\infty} f(x)\,dx = \frac{\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{y^2/2}\, dy = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = 1 //$$

ch. 3

* Mean & Variance

Mean $E(X) = \mu$

variance $v(X) = E(X^2) - \{E(X)\}^2 = \sigma^2$

* standard normal distribution

It is the special case of normal distribution, when $\mu = 0$ & $\sigma = 1$ the probability density function is given by,

$$g(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dy \quad , -\infty < z < \infty$$

& is denoted by

$$z = \frac{x - \mu}{\sigma}$$

let

$$z = \frac{x - \mu}{\sigma}$$

be the standard variable (Random variable)

Then, $z \sim N(0, 1)$

with probability density function

$$P(Z = z) = \frac{1}{2\pi} e^{-z^2/2}$$

is called the standard normal probability distribution of X. The corresponding distribution function is,

$$F(z) = P(Z = z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \quad , -\infty < z < \infty$$

$$= \phi(z)$$

8) Define standard normal distribution.
Write down its importance in engineering field.

Note:- $p(a \leq X \leq b) = p\left(\frac{a-\mu}{\sigma} \leq z \leq \frac{b-\mu}{\sigma}\right)$

$$= \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$

$$p(X \geq a) = 1 - p(X \leq a)$$

$$= 1 - \phi\left(\frac{a-\mu}{\sigma}\right)$$

Suppose,

$$P(X \leq b) = \phi\left(\frac{b-\mu}{\sigma}\right)$$

* Importance of Normal distribution.

The importance of Normal distribution are,

a) Most of distribution such as Binomial, Poisson, Hypergeometric distribution, etc. can be approximated using Normal distribution.

b) It is extensively used in large sample test to estimate parameters.

c) Many of the distribution can be transformed to normal distribution.

d) For large samples, the distribution of sample mean & variance follow normal distribution.

* Application of Normal Distribution.

- It is widely used in industries for quality control of raw materials.

- It is also used in analysis of ~~large sample test~~ manufactures items & their ability to meet specification.

Imp

* Area property of standard normal distribution.

The standard normal variable corresponding to x is z.

$$z = \frac{x-\mu}{\sigma}$$

when $x = \mu+\sigma$, $z = \frac{x-\mu}{\sigma} = \frac{\mu+\sigma-\mu}{\sigma} = 1$

$x = \mu-\sigma$, $z = \frac{x-\mu}{\sigma} = \frac{\mu-\sigma-\mu}{\sigma} = -1$

$x = \mu+2\sigma$, $z=2$

$x = \mu-2\sigma$, $z=-2$

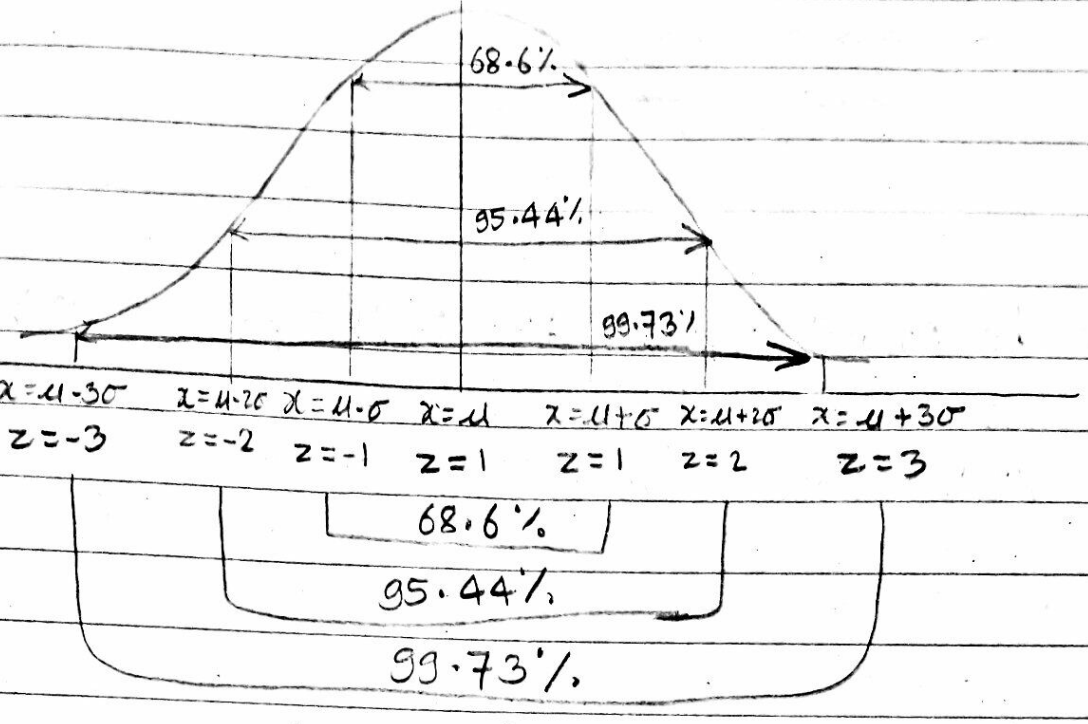
$x = \mu+3\sigma$, $z=3$

$x = \mu-3\sigma$, $z=-3$

$\vdots$

$x = (\mu \pm n\sigma)$, $z = \pm n$

ch. 3



Area property of normal probablity distribution.

Imp

* Normal approximation to the Binomial distribution

For very large value of n binomial distribution is not used as it is tedious so we ~~seed~~ used normal distribution to approximate the binomial distribution.

Thus, the normal distribution is a limiting case of the binomial distribution under the following condition.

i) when no. of trail 'n' is indefinetely large. i.e $n \to \infty$

ii) Neither 'p' nor 'q' is very small.

If X is a random variable having binomial distribution with parameter n & p, the limiting form of the distribution function is given by,

$$Z = \frac{X - np}{\sqrt{npq}} , \quad q = 1 - p$$

or, $n \to \infty$ then,

$$F(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \quad , -\infty < z < \infty$$

Imp

* The normal approximation to the poisson distribution.
Normal distribution is a limiting case of poisson distribution if average number (λ) is very large i.e $\lambda \to \infty$
so,

ch. 3

chaitra

Q) The breakdown voltage X of a randomly choosen diode of a particular type is known to be normally distributed with mean 40 volts & variance 2.25 volts. What is the probability that the breakdown voltage will be:

i) Between 39 volts and 42 volts

ii) Less than 44 volts

iii) More than 43 volts.

Sol^n Given, $X \sim N(40, 1.5^2)$, $\mu = 40$, $\sigma = 1.5$

i) $P(39 \le X \le 42) = P(a \le X \le b)$

$$= \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$

$$= \phi\left(\frac{42-40}{1.5}\right) - \phi\left(\frac{39-40}{1.5}\right)$$

$$= \phi(1.33333) - \phi(-0.66667)$$

$$= 0.9082 - 0.2514$$

$$= 0.6568.$$

ii) $P(X \le b) = P(X \le 44)$

$$= \phi\left(\frac{b-\mu}{\sigma}\right)$$

$$\cancel{= \phi\left(\frac{42-40}{1.5}\right)} \qquad = \phi\left(\frac{44-40}{1.5}\right)$$

$$\cancel{= \phi(1.33)} \qquad = \phi(2.67)$$

$$\cancel{= 0.9082} \qquad = 0.9962$$

iii) $P(43 \le X) = P(\cancel{43 <} a \le X)$

$= 1 - P(X \le a)$

$= 1 - \phi\left(\frac{a-\mu}{\sigma}\right)$

$= 1 - \phi\left(\frac{43-40}{1.5}\right)$

$= 1 - \phi(2)$

$= 1 - 0.9772$

$= 0.0228$

Q) The breakdown voltage X of randomly choosen diode of a particular type is known to be normally distributed with mean 40 & standard deviation 1.5 volts. What is the probablity that the breakdown voltage will be,

i) Between 39 & 42 volts

ii) At most 43 volts

iii) At least 39 volts.

soln Given, $X \sim N(\mu, \sigma^2)$

i.e. $X \sim N(40, 1.5)^2$, $\mu = 40$, $\sigma = 1.5$

i) $P(39 \le X \le 42) = P(a \le X \le b)$

$= \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$

$= \phi\left(\frac{42-40}{1.5}\right) - \phi\left(\frac{39-40}{1.5}\right)$

$= \phi(1.33) - \phi(-0.67)$
$= 0.9082 - 0.2514$
$= \underline{\underline{0.6568}}.$

ii) $P(X \le 43) = P(X \le b)$
$= \phi\left(\frac{b-\mu}{\sigma}\right)$
$= \phi\left(\frac{43-40}{1.5}\right)$
$= \phi(2)$
$= \underline{\underline{0.9772}}$

iii) $P(39 \le X) = 1 - P(X \le a)$
$= 1 - P(X \le 39)$
$= 1 - \phi\left(\frac{a-\mu}{\sigma}\right)$
$= 1 - \phi\left(\frac{39-40}{1.5}\right)$
$= 1 - \phi(-0.67)$
$= 1 - 0.2514$
$= \underline{\underline{0.7486}}$

Q) The marks obtained by IOE students in statistics are 50 on average with variance 16. If 5000 students have given the exam, find the following.
a) The number of students securing marks less than 40?

b) The number of students securing marks between 35 to 60?

sol? Given ~~N~~ $X \sim N(\mu, \sigma^2) = X \sim N(50, 16^2)$; $\mu = 50, \sigma = 16$

Now,

a) ⇒ no. of students securing marks less than 40?

$$p(X \leq 40) = P(X \leq b)$$
$$= \phi\left(\frac{b-\mu}{\sigma}\right)$$
$$= \phi\left(\frac{40-50}{16}\right)$$
$$= \phi(-0.63)$$
$$= 0.2643$$

∴ no. of students $= 5000 \times 0.2643 = 1321.5 \approx 1322$

~~b)~~

b) ⇒ no. of students securing marks between 35 to 60?

$$p(35 \leq X \leq 60) = p(a \leq X \leq b)$$
$$= \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$
$$= \phi\left(\frac{60-50}{16}\right) - \phi\left(\frac{35-50}{16}\right)$$
$$= \phi(0.63) - \phi(-0.94)$$
$$= 0.7357 - 0.1736$$
$$= 0.5621$$

so, no. of students $= 0.5621 \times 5000$

$= 2810.5 \approx 2811$

ch 3

8) The mean inside diameter of a sample of 200 washers produced by a machine is 0.502 cm and the standard deviation is 0.005 cm. The purpose for these washers are interned allows a maximum tolerance in the diameter of 0.496 to 0.508 cm, otherwise the washers are considered defective. Determine the percentage of defective washers produced by the machine. Assume the diameter is normally distributed.

sol^n Given $X \sim N(\mu, \sigma^2) = X \sim N(0.502, 0.005^2)$

i.e $\mu = 0.502$ & $\sigma = 0.005$

no. of washers = 200

Now,

probablity of getting diameter between 0.496 to 0.508 cm is given as,

$$P(a \le x \le b) = P(0.496 \le x \le 0.508)$$

$$= \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$

$$= \phi\left(\frac{0.508-0.502}{0.005}\right) - \phi\left(\frac{0.496-0.502}{0.005}\right)$$

$$= \phi(1.2) - \phi(-1.2)$$

$$= 0.8849 - 0.1151$$

$$= 0.7698$$

So, probablity of getting defective machine $= 1 - 0.7698$

$= 0.2302$

& hence percentage of getting defective washers is 23.02%

ch. 3

070 chaitra

S) The time required to assemble a piece of machinery is a random variable having approximately a normal distribution with mean 12.9 minutes and standard deviation of 2 minutes. What are the probablities that the assembly of a piece of machinery of ~~the~~ this kind will take a) at least 11.5 minutes b) between 11.0 to 14.8 minutes?

Sol^n Given $X \sim N(\mu, \sigma^2) = X \sim N(12.9, 2^2)$

i.e $\mu = 12.9$ & $\sigma = 2$.

a) at least 11.5 minutes

$$p(11.5 \leq X) = 1 - p(X \leq 11.5)$$

$$= 1 - \phi\left(\frac{b-\mu}{\sigma}\right)$$

$$= 1 - \phi\left(\frac{11.5 - 12.9}{2}\right)$$

$$= 1 - \phi(-0.70)$$

$$= 1 - 0.2420$$

$$= 0.7580$$

b) between 11.0 to 14.8 minutes

$$p(11.0 \leq X \leq 14.8) = p(a \leq X \leq b)$$

$$= \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$

$$= \phi\left(\frac{14.8 - 12.9}{2}\right) - \phi\left(\frac{11.0 - 12.9}{2}\right)$$

$$= \phi(0.95) - \phi(-0.95)$$

$= 0.8289 - 0.1711$
$= \underline{\underline{0.6578}}$

5.

3.3) Gama distribution

A continuous random variable X is said to have gamma distributions with parameters α & β, if X has the probablity density function (p.d.f)

$$p(X=x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha} \Gamma\alpha}, \quad 0 < x < \infty, \quad \alpha > 0, \beta > 0$$

& distribution function

$$F(x) = p(X=x) = \int_{-\infty}^{x} \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha} \Gamma\alpha} \, dx$$

The mean & variance of this distribution is given as.

Mean $(\mu) = E(X) = \alpha\beta$ &
variance $(\sigma^2) = V(X) = \alpha\beta^2$

* Application of gamma distribution

- This distribution is useful in the study of the length of life of industrial equipments, electrical supply in certain areas, distribution of petrol & in other yields.

ch.3

071 chaitra

Q) The daily consumption of electric power in a certain city follow a gamma distribution with $\alpha = 2$ & $\beta = 3$. If the power plant of this city has daily capacity of 12 million Kilowatt hours, what is the probablity that this power supply will be inadequate on any given day?

Sol^n let X = random variable associated with consumption of electricity

$\alpha = 2$ & $\beta = 3$

Now,

$$f(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha} \Gamma\alpha}, \quad 0 < x < \infty$$

$$= \frac{x^{2-1} e^{-x/3}}{3^2 \Gamma 2}$$

$$= \frac{x \cdot e^{-x/3}}{9}$$

So, $\cancel{P(X \leq 12)} = P(X > 12)$

$$= 1 - P(X \leq 12)$$

$$= 1 - \int_{0}^{12} \frac{x \cdot e^{-x/3}}{9} dx$$

$$= 1 - \frac{1}{9} \int_{0}^{12} x e^{-x/3} dx$$

$$\cancel{= 1 - \frac{1}{9} \left[x \right.}$$

from calculator

~~0.9915~~ 0.09158

ch. 3

3.4) chi-square distribution.

question no. ~~6~~ 7 & 8.

chapter: 4 (Sampling Distribution)

4.1) Population & sample.

*Population

Population is a group of entire item (or individual of interest) under the investigation. There are two types of population.

a) Target population

b) Sampling population

a) Target population

It is the population for which representative information is desired.

b) Sampling population

It is the population from which sample is taken.

For example.

The no. of people in letangbhogateni municipality is population, the no. of people speaking newari language may be target population and if certain no. of people are selected from the population for example 50, to find no. of people speaking newari is sample.

*Sample

Some selected items from the population is called sample and the process of selecting sample from the population in order to draw conclusion about population is called sampling.

Parameter

A parameter is summary measure that describe the characteristics of a population. Population mean (μ), population standard deviation (σ), population correlation coefficient (ρ), population number (N), are the examples of parameter.

statistic

A statistic is a summary measure ~~of the sample which is used to estimate the parameter of~~ that describe the characteristics of sample. Sample mean ($\bar{x}$), sample standard deviation (s), sample correlation coefficient (r), sample numbers (n), are the examples of statistic.

Type of sample

i) Finite sample

ii) Infinite sample

i) Finite sample

Let a set of observation $x_1, x_2, x_3 \ldots x_n$ from a random sample of size 'n' from population of finite number N. If the values are choosen such that subset of 'n' of the N elements of the population has the same probablity of being selected.

ii) Infinite sample

If a set of observation $x_1, x_2, x_3 \ldots x_n$ from a random

sample of size 'n' from the infinite population $f(x)$ if,

i) Each x_i is a random whose distribution is given by $f(x)$,

ii) These n random variables are independent.

* Difference between population & sample.

The difference between population & sample are given below.

Population	Sample
1) It is collection of item which is to be considered. Example People in Nepal.	1) It is a part or portion of population considered for purpose of study. For example people speaking Newari in Nepal.
2) Population is defined by parameter.	2) Sample is defined by statistics.
3) The symbols used in population are, population size $= N$ population mean $= \mu$ " standard deviation $= \sigma$ " variance $= \sigma^2$ " proportion $= P$ " correlation coefficient $= \rho$	3) The symbols used in sample are, sample size $= n$ sample mean $= \bar{x}$ " standard deviation $= s$ " variance $= s^2$ " proportion $= \hat{p}$ " correlation coefficient $= r$

8) Define the central limit theorem. Write its applications.

4.2) Central limit theorems

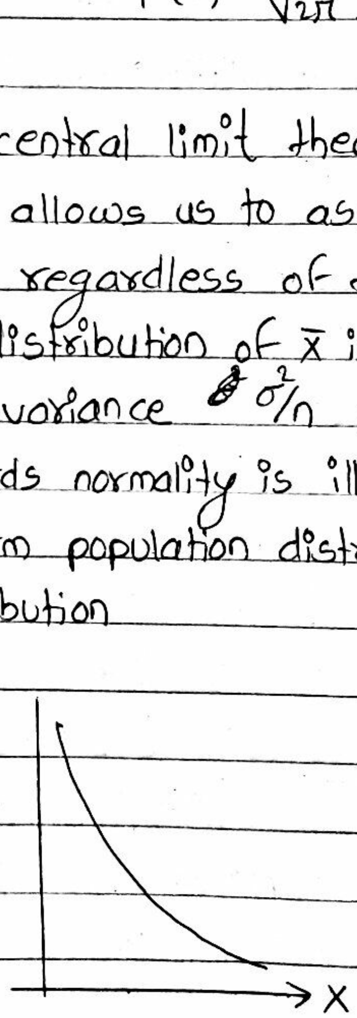
If $\bar{x}$ is the mean of sample of size 'n' taken from a population having the mean 'μ' & finite varience σ^2, then

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

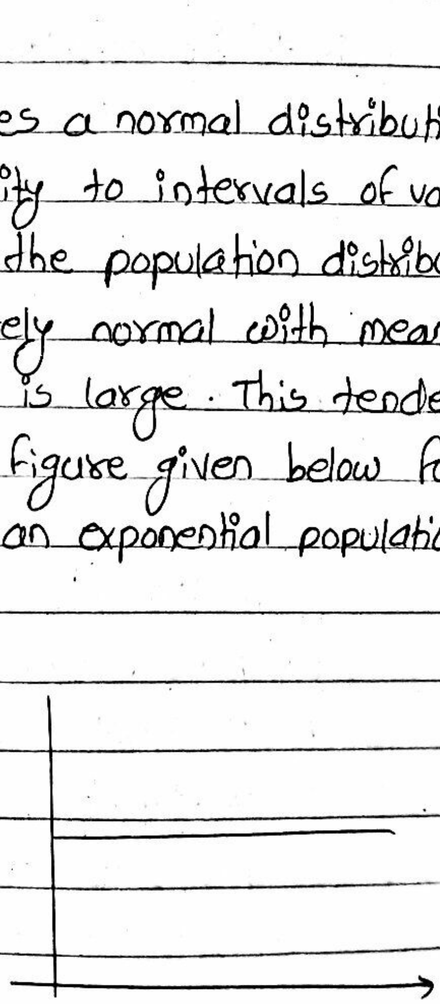
is a random variable distribution function approches that of the standard normal distribution as $n \to \infty$

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} \, dt$$

The central limit theorem provides a normal distribution that allows us to assign probablity to intervals of values of $\bar{x}$ regardless of the form of the population distribution, the distribution of $\bar{x}$ is approximately normal with mean μ and variance σ^2/n whenever n is large. This tendency towards normality is illustrated in figure given below for a uniform population distribution and an exponential population distribution



Exponential population distribution



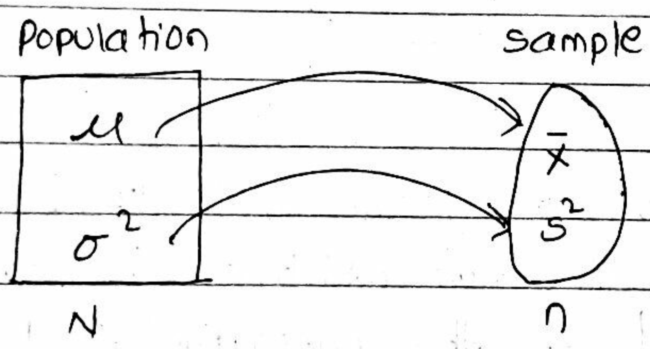
uniform population distribution

Q) what do you mean by sampling distribution of a sample mean & its standard error? Explains with example.

4.3) Sampling distribution of sample mean

let $x_1, x_2, x_3, \ldots x_n$ denote the independent identically distributed random variable of sample of size n from a population of size N having mean μ & finite variance σ^2. Then the sample mean $\bar{x} = \frac{\Sigma x_i}{n}$ is a random variable with mean μ & variance (σ^2/n). The probablity distribution of the random variable $\bar{x}$ is called the sampling distribution of the mean. Then $\bar{x} \sim N(\mu, \sigma^2/n)$

$$\text{If } n>30;\ z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$



$$\bar{x} = \frac{\Sigma x_i}{n}, \quad s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1}$$

$$= \frac{n\Sigma x_i^2 - (\Sigma x_i)^2}{n(n-1)}$$

Rule of Thumb

If $n > 30$, the central limit theorem can be used.

4.4) Sampling distribution of sampling proportion.

Let a population be binomially distributed with $P = X/n$ = propability of success & $q = 1 - P$, the probablity of failure. Now, consider all possible samples of size n drawn from this population. Let $\bar{p}$ be the proportion $\bar{p}$ with mean '$\mu_{\bar{p}}$' & variance $(\sigma_{\bar{p}}^2)$ given by $\mu_{\bar{p}} = P$ & $\sigma_{\bar{p}}^2 = \frac{\bar{p}\bar{q}}{n}$

If $n > 30$ & both n_p & $n_q \geq 15$, then,

$$z = \frac{\bar{p} - P}{\sqrt{\frac{\bar{p}\bar{q}}{n}}} \sim N(0,1)$$

obg chaitra

Q) A population consists of 5, 6, 9, 12. Consider all possible samples of size two which can be drawn without replacement from this population. Find,

i) population mean & population standard deviation
ii) mean of sampling distribution of mean
iii) standard error of sampling distribution of mean.

Soln Given, population values $(X) = 5, 6, 9, 12$ i.e $N = 4$
possible sample $(n) = 2$

a) ⇒ population mean $(\mu) = \frac{\Sigma X}{N}$

$$= \frac{5+6+9+12}{4} = \frac{32}{4} = 8$$

population standard deviation $(\sigma) = \sqrt{\frac{\Sigma(X-\mu)^2}{N}}$

$$= \sqrt{\frac{(5-8)^2+(6-8)^2+(9-8)^2+(12-8)^2}{4}} = \sqrt{\frac{9+4+1+16}{4}}$$

$$= \sqrt{\frac{30}{4}} = 2.74.$$

b) ⇛ The no. of samples are ${}^{N}C_{n} = {}^{4}C_{2} = 6$ without replacement & the calculation is given below,

Sample no.	Sample values	$\bar{x}$	s^2	s
1	5,6	5.5		
2	5,9	7.0		
3	5,12	8.5		
4	6,9	7.5		
5	6,12	9.0		
6	9,12	10.5		
		48		

The mean of sampling distribution of mean,

$$= \mu_{\bar{x}} = \frac{\Sigma \bar{x}_i}{m} = \frac{5.5+7.0+8.5+7.5+9.0+10.5}{6}$$

$$= \frac{48}{6} = 8$$

Here, sample mean = population mean,

c) ⇛ We know, standard error on sampling without replacement

$$S_{\bar{x}} = E(\bar{x}) = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{2.74}{\sqrt{2}} \sqrt{\frac{4-2}{4-1}} = 1.58$$

08 chaitra

Q) State the central limit theorem. A random sample of size 100 is taken from an infinite population having the mean 76 & variance 256. What is the probablity that the sample mean will be between 75 & 78?

Paxadi garya xa.

Soln Numerical part,

Given, no. of sample $(n) = 100$

mean $(\mu) = 76$

variance $(\sigma^2) = 256$

standard deviation $= \dfrac{\sigma}{\sqrt{n}} = \dfrac{16}{\sqrt{1000}} = 1.6$

$$\therefore P(75 \le \bar{x} \le 78) = P\left(\frac{a-\mu}{\sigma/\sqrt{n}} \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le \frac{b-\mu}{\sigma/\sqrt{n}}\right)$$

$$= P\left(\frac{75-76}{1.6} \le \frac{\bar{x}-76}{1.6} \le \frac{78-76}{1.6}\right)$$

$$= P(-0.6 \le z \le 1.25)$$

$$= P(0 \le z \le 0.60) + P(0 \le z \le 1.25)$$

$$= 0.2257 + 0.3944$$

$$= 0.6201$$

Q) Define the central limit theorem. A sample of 100 mobile battery cells tested to find the length of the life produced the following results as mean 13 months & standard deviation of 3 months. Assuming the data to be normally distributed by using central limit theorem what percentage

of battery cells expected to have average life?
i) more than 15 months ii) less than 9 months.

Sol^n Given, no. of sample $(n) = 100$

mean $(\mu) = 13$

standard deviation $(\sigma) = 3$

Since $n > 30$ so using central limit theorem

so, ~~standard~~ $\frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{100}} = \frac{3}{10} = 0.3$

Now,

i) more than 15 months

$$p(\bar{x} > 15) = p\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{15-13}{\sigma/\sqrt{n}}\right)$$

$$= p\left(z > \frac{2}{0.3}\right)$$

$$= p(z > 6.67)$$

$$= \cancel{p(} \; p(z < -6.67)$$

$$= 1 - p(z < 6.67)$$

$$= 1 -$$

071 chaitra

8) state central limit theorem. An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed with mean equal to 8000 hours and standard deviation of 4 hours. Find the probablity that the a random sample of 16 bulbs will have an average life of less than 12775 hours.

soln For numerical part,

no. of sample $(n) = 16$

mean $(\mu) = 8000$ hours

standard deviation $(\sigma) = 4$ hours

$p(\bar{x} < 12775) = ?$

Now,

using central limit theorem

$$p(\bar{x} < 12775) = p\left(z < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right)$$

$$= p\left(z < \frac{12775 - 8000}{4/\sqrt{16}}\right)$$

$$= p(z <$$

chapter:-5, correlation & regression. Q 9 & 10

5.1) correlation

correlation is the degree and direction of association between two (or more) variables.

* Types of correlation

i) simple correlation
ii) Partial correlation
iii) Multiple correlation

i) Simple correlation

The correlation between two variables is called simple correlation.

For example.

→ Height & weight of students
→ Income & expenditure of families
→ Amount of rainfall & production of crops
→ C.P & S.P of products, etc.

ii) Partial correlation

In this correlation common factor of the two variable are eliminated. let x, y, z be three variables, then partial correlation between x & y is denoted by $r_{xy.z}$ & given by

$r_{xy.z}$ = Partial correlation between x & y keeping z constant. (અગ્રે વાંચી પહાડી છ

ch. 5

$$= \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{1-r_{xz}^2} \cdot \sqrt{1-r_{yz}^2}}$$

iii) Multiple correlation

Let x, y, z be the three variable then multiple correlation co-efficient between x & y is denoted by $r_{z.xy}$ & given as,

$r_{z,xy}$ = Multiple correlation coefficient between dependent variable z & joint effect of independent variable x & y on z.

$$\text{or, } R_{z.xy} = \sqrt{\frac{r_{xz}^2 + r_{yz}^2 - 2 r_{xy} r_{xz} r_{yz}}{1 - r_{xy}^2}}$$ (અઝીં નામો પહાડી છે)

5.2) Least square methods

We know general equation

$$y = a + bx \quad \cdots \quad (i)$$

where, a & b are constant

Normalized equations are

$$\Sigma y = na + b\Sigma x \quad \text{———} \quad (ii)$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad \text{———} \quad (iii)$$

where,

Σx, Σy, Σxy & Σx^2 are determined from given data.

Karl pearson's coefficient (r) = $\dfrac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2} \cdot \sqrt{n\Sigma y^2 - (\Sigma y)^2}}$ ch. 5

* Coefficient of correlation

Karl Pearson's correlation co-efficient between two variable 'x' & 'y' is defined as,

$$r = \frac{Cov(X,Y)}{\sqrt{v(X)}\sqrt{v(Y)}}$$

where,

r = correlation co-efficient

$cov(X,Y)$ = co-variance between x & y

$v(X)$ = variance of x

$v(Y)$ = variance of y

$\therefore$ Correlation co-efficient 'r' can be written as,

i) $$r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2} \cdot \sqrt{\Sigma(Y-\bar{Y})^2}}$$

ii) $$r = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{n\Sigma X^2 - (\Sigma X)^2} \cdot \sqrt{n\Sigma Y^2 - (\Sigma Y)^2}}$$

Let $U = X - A$

$V = Y - B$

where,

A = Assumed mean of X

B = Assumed mean of Y

$$\therefore r = \frac{n\Sigma UV - (\Sigma U)(\Sigma V)}{\sqrt{n\Sigma U^2 - (\Sigma U)^2} \cdot \sqrt{n\Sigma V^2 - (\Sigma V)^2}}$$

For. Bivariate frequency distribution

$$r = r_{xy} = \frac{\Sigma fxy - \frac{1}{N}(\Sigma fx)(\Sigma fy)}{\sqrt{\Sigma fx^2 - \frac{1}{N}(\Sigma fx)^2} \cdot \sqrt{\Sigma fy^2 - \frac{1}{N}(\Sigma fy)^2}}$$

contd અગાઉનો ii) <u>Partial correlation</u>

The partial correlation coefficient may be defined as a measure of degree of relationship between any two variables out of a set of variables eliminating the common association of remaining variables with both of them.

The partial correlation coefficient between the variables x_1 & x_2 avoiding the effect of x_3 is defined as,

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{13}^2} \cdot \sqrt{1-r_{23}^2}}$$

similarly,

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{23}^2}}$$

& $$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{1-r_{12}^2} \cdot \sqrt{1-r_{13}^2}}$$

$\boxed{\text{properties}}$

i) $r_{12} = r_{21}$

ii) $-1 \le r_{12.3} \le 1$

iii) Multiple correlation

The multiple correlation coefficient ~~on the variable x_1 with joint effect of x_2 & x_3~~ studies the relationship between the dependent variables & joint effect of independent variables.

The multiple correlation coefficient on the variable x_1 with joint effect of x_2 & x_3 is

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}$$

similarly, $R_{2.13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{13}^2}}$

$$R_{3.12} = \sqrt{\frac{r_{13}^2 + r_{23}^2 - 2r_{12}r_{23}r_{13}}{1 - r_{12}^2}}$$

$\boxed{\text{Properties}}$

i) $R_{1.23} = R_{1.32}$ ... etc

ii) $0 \leq R_{1.23} \leq 1$

ch. 05

* Regression

Regression is a mathematical measurement of average relationship between variable in terms of origional unit of data. In regression analysis, there are two types of variable i.e. independent variable & dependent variable.

i) Dependent variable

Those variable whose values is to be estimated is called dependent variable.

ii) Independent variable

Those variable ~~whose~~ which are used for prediction is called independent variable.

Q) Consider following sample result, where the number of data point 'x' is used to predict computer processing time 'y' (in sec).

x	105	511	401	622	330	211	332	332
y	44	214	193	299	143	112	155	131

Use the method of least square to determine the expression for the estimated regression line. The number of data point is 200.

Sol^n We know,

Equations of straight line.

ch. 5

$y = a + bx$ —— (i)

Normal equations are,

$\Sigma y = na + b\Sigma x$ —— (ii)

$\Sigma xy = a\Sigma x + b\Sigma x^2$ —— (iii)

& from the given table & calculators,

$\Sigma y = 1291$

$\Sigma x = 2844$

$\Sigma xy = 543119$

$\Sigma x^2 = 1193700$

& $n = 8$

substituting above values in equations (ii) & (iii) we get,

$1291 = 8a + \cancel{2488b}\ 2844b$

i.e $8a + \overset{2844}{\cancel{2488}}b = 1291$ —— (a)

& $543119 = 2844a + 1193700b$

i.e $2844a + 1193700b = 543119$ —— (b)

solving equations (a) & (b) we get

$a = -2.44$ &

$b = 0.46$

substituting values of a & b in equation (i) we get.

$y = -2.44 + 0.46x$

$= -2.44 + 0.46 \times 200$

$\therefore y = \underline{89.56 \text{ sec.}}$

ch. 5

072 chaitra.

Q) An artical in wear (Vol 152, 1992, PP. 171-181) presents data on the fretting wear of mind steel & oil viscasity. Representative data follow, with x = oil viscosity & y = water volume $(10^{-4} mm^3)$.

y	240	181	193	155	172	110	113	75	94
x	1.6	9.4	15.5	20.0	22.0	35.5	43.0	40.5	33.0

i) Fit the sample linear regression model using least
ii) Predict fretting wear when viscosity, $x = 30$.

Soln We know,

Equation of straight line

$$y = a + bx \quad \text{—— (i)}$$

& normal equations are

$$\left.\begin{aligned} \Sigma y &= na + b\Sigma x \\ \& \Sigma xy &= a\Sigma x + b\Sigma x^2 \end{aligned}\right\} \text{—— (ii)}$$

~~Fo~~ From the given data table, using calculators we find,

$\Sigma y = 1333$

$\Sigma x = 220.5$

$\Sigma xy = 26864.4$

$\Sigma x^2 = 7053.67$ & $n = 9$

Substituting the above values in equation (ii) we get

$$1333 = 9a + 220.5b$$

i.e. $9a + 220.5b = 1333$ —— (iii)

& $26864.4 = 220.5a + 7053.67b$

i.e $220.5a + 7053.67b = 26864.4$ —— (iv)

Solving equations (iii) & (iv) we get,

y on x ⇒ $y = a + bx$

ch.5

$a = 234.07$ &

$b = -3.51$

substituting the values of 'a' & 'b' in equation (i) we get

i) ⇒ $y = 234.07 - 3.51x$

ii) ⇒ when viscosity, $x = 30$

$y = 234.07 - 3.51 \times 30$

or $y = 128.77$

i.e volume, $y = 128.77 \times 10^{-4} mm^3$

071 chaitra
071 shrawn

Q) The following data gives the number of twists required to break a certain kind of forged alloy bar & percentage of alloying element A present in the metal.

No. of twists	41	49	69	65	40	150	58	57	31	36
% of element A	10	12	14	15	13	12	13	14	13	12

i) Fit the regression equation of number of twists on percentage of element A. Determine the predicted no. of twists required to break the alloy when percentage of element is 20.

ii) Find 99% confidence interval for the regression coefficient (i.e. slope).

Soln We know,

Equation of straight line

$y = a + bx$ ——— (i)

& normal equations are,

$$\left.\begin{aligned} \Sigma y &= na + b\Sigma x \\ \& \ \Sigma xy &= a\Sigma x + b\Sigma x^2 \end{aligned}\right\} \text{—— (ii)}$$

let No. of twists be y & percentage of element A be x

Then, from the given table & calculations, we get

$\Sigma y = 496$

$\Sigma x = 128$

$\Sigma xy = 6446$

$\Sigma x^2 = 1656$

$n = 10$

substituting above values in equation (ii) we get.

$496 = 10a + 128b$

i.e $10a + 128b = 496$ —— (iii)

& $6446 = 128a + 1656b$

i.e $128a + 1656b = 6446$ —— (iv)

Solving equations (iii) & (iv) we get

$a = -21.09$ &

$b = 5.52$

substituting the values of a & b in equation (i) we get.

i) ⇒ $y = -21.09 + 5.52x$

Also, if $x = 20\%$.

then, $y = -21.09 + 5.52 \times 20$

i.e $y = 89.31$

since no. are not in decimal

so, $89.31 \approx \underline{\underline{89}}.$

ch. 5

ii) ⇒

070 chaitra

Q) Observation on the yield of a chemical reaction taken at various temperature was recorded as follows,

X (°c)	150	150	200	250	250	300	150
Y %	75.4	81.2	85.5	89	90.5	96.7	75.4

Fit a simple linear regression & estimate value of yield at 200°C.

Sol^n We have,

Equation of straight line

$$y = a + bx \quad \text{--- (i)}$$

& normal equations are,

$$\left.\begin{aligned} \Sigma y &= na + b\Sigma x \\ \& \ \Sigma xy &= a\Sigma x + b\Sigma x^2 \end{aligned}\right\} \text{--- (ii)}$$

From the given table & calculation we get

$$\Sigma x = 1450$$

$$\Sigma y = 593.7$$

$$\Sigma xy = 125785$$

$\Sigma x^2 = 322500$

$\& \ n = 7$

substituting above values in equation (ii) we get.

$$593.7 = 7a + 1450b$$

$$\text{i.e } 7a + 1450b = 593.7 \quad \text{—— (iii)}$$

$$\& \ 125785 = 1450a + 322500b$$

$$\text{i.e } 1450a + 322500b = 125785 \quad \text{—— (iv)}$$

solving equations (iii) & (iv) we get

$$a = 58.58$$

$$\& \ b = 0.13$$

Now, At $200^{\circ}C$

$$y = a + bx$$

$$\text{or, } y = 58.58 + 0.13 \times 200$$

$$\text{i.e } Y = 84.58\ \%$$

069 chaitra

Q) The simple correlation coefficient between fertilizer (X_1), seeds (X_2) and productivity (X_3) are $r_{12} = 0.69$, $r_{13} = 0.64$ & $r_{23} = 0.85$. calculate the partial correlation $r_{12.3}$ & multiple correlations $R_{1.23}$.

Sol^n Given, $r_{12} = 0.69$, $r_{13} = 0.64$, $r_{23} = 0.85$

$$r_{12.3} = ? \ \& \ R_{1.23} = ?$$

Now, we have,

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1 - r_{13}^2} \cdot \sqrt{1 - r_{23}^2}}$$

$$= \frac{0.69 - 0.64 \times 0.85}{\sqrt{1-(0.64)^2} \cdot \sqrt{1-(0.85)^2}}$$

$$\therefore r_{12.3} = 0.36$$

$$\& \; R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2 r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{(0.69)^2 + (0.64)^2 - 2 \times 0.69 \times 0.64 \times 0.85}{1-(0.85)^2}}$$

$$\therefore R_{1.23} = 0.6974$$

069 chaitra Q) An article in Concrete Research presented data on compressive strength X and intrinsic permeability Y of various concrete mixes & cures. Summary quantities are $n = 14$, $\Sigma y = 572$, $\Sigma y^2 = 23,530$, $\Sigma x = 43$, $\Sigma x^2 = 157.42$ & $\Sigma xy = 1697.80$. Assume that the two variables are related according to the simple linear regression model.

i) Calculate the least squares estimates of the slope & intercept.

ii) Use the equation of the fitted line to predict what permeability would be observed when the compressive strength is $x = 4.3$.

Soln Given,

$n = 14,\ \Sigma y = 572,\ \Sigma y^2 = 23{,}530,\ \Sigma x = 43,\ \Sigma x^2 = 157.42$ &
$\Sigma xy = 1697.80.$

Now, we have equation of straight line

$$y = a + bx \quad \text{—(i)}$$

& normal equations are

$$\left.\begin{aligned} \Sigma y &= na + b\Sigma x \\ \& \ \Sigma xy &= a\Sigma x + b\Sigma x^2 \end{aligned}\right\} \text{——(ii)}$$

Putting the given values in equation (ii) we get,

$$572 = 14a + 43b \quad \text{——(iii)}$$

$$\& \ 1697.80 = 43a + 157.42b \quad \text{——(iv)}$$

Solving equations (iii) & (iv) we get

$$a = 48.01$$

$$\& \ b = -2.3298$$

i) ⇒ slope (m) = b = ~~2.30~~ −2.3298

intercept (c) = a = 48.01

ii) ⇒ when the compressive strength $x = 4.3$

$$y = a + bx$$

$$= 48.01 - 2.3298 \times 4.3$$

$$\therefore y = 37.99186$$

i.e Intrinsic permeability (y) = 37.99186

ch. 5

068 chaitra Q) The Following table gives the age of the cars of a certain company and annual maintenance costs.

Age of cars (Years) (x)	2	4	6	8	10
Maintenance costs (Rs.) (y)	10	15	22	32	46

obtain the regression ~~coefficient~~ equation for cost related to age & also estimate the cost of maintance for 10 yrs old car.

Soln We have

Equation of straight line

$$y = a + bx \quad \text{—— (i)}$$

& normal equations are

$$\left.\begin{aligned} \Sigma y &= na + b\Sigma x \\ \& \ \Sigma xy &= a\Sigma x + b\Sigma x^2 \end{aligned}\right\} \text{—— (ii)}$$

From the given table & calculations we have

$$\Sigma y = 125$$
$$\Sigma x = 30$$
$$\Sigma xy = 928$$
$$\Sigma x^2 = 220$$
$$n = 5$$

substituting above values in equation (ii) we get

$$125 = 5a + 30b \quad \text{—— (iii)}$$

$$\& \ 928 = 30a + 220b \quad \text{—— (iv)}$$

solving equations (iii) & (iv) we get

$$a = -1.7 \ \&$$
$$b = 4.45$$

ch. 5

substituting values of a & b in equation (i) we get

$$y = -1.7 + 4.45x$$

Hence, the cost for maintanance of 10 years old car is, $y = -1.7 + 4.45 \times 10$

$$= 12.8$$

i.e Rs. 12.8.

068 chaitra Q) The simple correlation coefficient between temperature (X_1), corn yield (X_2) & rainfall (X_3) are $r_{12} = 0.59$, $r_{13} = 0.46$ & $r_{23} = 0.77$. Calculate the partial correlation coefficient $r_{12.3}$ & multiple correlation $R_{1.23}$.

Soln Given,

$r_{12} = 0.59$

$r_{13} = 0.46$

$r_{23} = 0.77$

Now, we have,

partial correlation coefficient $r_{12.3} = \dfrac{r_{12} - r_{13} r_{23}}{\sqrt{1-r_{13}^2} \cdot \sqrt{1-r_{23}^2}}$

or, $r_{12.3} = \dfrac{0.59 - 0.46 \times 0.77}{\sqrt{1-(0.46)^2} \cdot \sqrt{1-(0.77)^2}}$

$\therefore r_{12.3} = 0.416$

Again,

ch. 5

~~mult~~ multiple correlation coefficient $(R_{1.23}) = \sqrt{\dfrac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}$

or, $R_{1.23} = \sqrt{\dfrac{(0.59)^2 + (0.46)^2 - 2 \times 0.59 \times 0.46 \times 0.77}{1 - (0.77)^2}}$

$\therefore R_{1.23} = \underline{\underline{0.59}}$

Q) In some determination of volume V of carbondioxide dissolved in a given. Volume of water of different temperature θ the following pair of the value were obtained.

θ	0	5	10	15
V	1.8	1.45	1.18	1.00

obtain by the method of least squares relation of the form $v = a + b\theta$ which best fit of these observation (2063 Ashadh)

Soln Given,

$V = a + b\theta$ —— ①

The normalized equations of ① are

$\Sigma V = na + b\Sigma\theta$ —— ⑪ &

$\Sigma V\theta = a\Sigma\theta + b\Sigma\theta^2$ —— ⑪⑪

From the table by the use of calculator we get.

$\Sigma V = 5.43$

$\Sigma\theta = 30$

$\Sigma V\theta = 34.05$

$\Sigma\theta^2 = 350$ & $n = 4$

substituting these values in normalized equation.

$$5.43 = 4a + 30b \quad \text{—— (IV)}$$

$$34.05 = 30a + 350b \quad \text{—— (V)}$$

solving equations (IV) & (V) we get.

$$a = 1.758$$

$$\& \; b = -0.0534$$

Hence, required equation is,

$$V = 1.758 - 0.0534\theta$$

Q) The following are the measurement of the air velocity & evaporation co-efficient of burning fuel droplets in an impulse engine.

x (cm/sec)	20	60	100	140	180	220	260	300	340	380
y (mm²/sec)	0.18	0.37	0.35	0.78	0.56	0.75	1.18	1.36	1.17	1.65

Fit a straight line to these data by the method of least square and use it to estimate the evaporation co-efficient of a droplet when the air velocity is 190 cm/sec.

soln The equation of straight line is,

$$y = a + bx \quad \text{—— (I)}$$

The normalized equations of (I) are

$$\Sigma y = na + b\Sigma x \quad \text{—— (II)}$$

$$\& \; \Sigma xy = a\Sigma x + b\Sigma x^2 \quad \text{—— (III)}$$

From the table using calculator, we get

$$\Sigma y = 8.35$$

$$\Sigma x = 2000$$

ch. 5

$\Sigma xy = 2175.4$

$\Sigma x^2 = 532000$

& $n = 10$

Substituting these values in above equations, we get,

$8.35 = 10a + 2000b$ — (IV)

$2175.4 = 2000a + 532000b$ — (V)

Solving equations (IV) & (V) we get,

~~$a = 0.069$~~

~~$b = 0.00383$~~

$a = \frac{457}{6600}$ & $b = 0.00383$

substituting the values of a & b in (1) we get

$y = \frac{457}{6600} + 0.00383x$

For $x = 190$ cm/sec

we have, evaporation coefficient $(y) =$

$\frac{457}{6600} + 0.00383 \times 190$

i.e $y = 0.7969$ mm²/sec.

062 Bhadra

Q) calculate the Karl person's coefficient of correlation between age & playing habits from the data given below.

Age	20	21	22	23	24	25
No. of students	500	400	300	240	200	160
Regular players	400	300	180	96	60	24

ch. 5

Q) The following table ~~are~~ gives age & percentage of blindness in respective age interval. Find out if there is any correlation between age and blindness.

Age (yrs)	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
% of blind	70	63	21	26	45	31	46	30

Soln Rewriting the table.

mid age (x)	5	15	25	35	45	55	65	75
% of blind (y)	70	63	21	26	45	31	46	30

By the use of calculators, from above table we get

$\Sigma x = 320$ $\quad\quad \Sigma x^2 = 17000$

$\Sigma y = 332$ $\quad\quad \Sigma y^2 = 15988$

$\Sigma xy = 11700$ $\quad\quad$ & $n = 8$

Now, correlation coefficient $(r) = \dfrac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$

$$\text{or, } r = \frac{8 \times 11700 - 320 \times 332}{\sqrt{8 \times 17000 - (320)^2}\sqrt{8 \times 15988 - (332)^2}}$$

$\therefore r = -0.00058 \approx 0$

Hence, Age & Blindness are weakly correlative with each other.

$x = 80$ 53616

$y = 75$ 54929

54190

ch. 5

2068 magh

Q) The following table shows the respective heights x & y of a sample of 11 fathers & their sons.

a) construct the scatters diagram

b) Find the squares regression line of y on x.

c) Find the least square regression line of x on y.

Ht. of Fathers(x)	65	63	67	64	68	62	70	66	68	67	60
Ht. of sons(y)	68	66	68	65	69	66	68	65	71	62	68

Soln: a) The scatter diagram is given as,

(In graph)

b) The square regression line of y on x is given as,

$$y = a + bx \quad \text{--- (I)}$$

& normalized equations of (I) are

$$\Sigma y = na + b\Sigma x \quad \text{--- (II)}$$

$$\& \ \Sigma xy = a\Sigma x + b\Sigma x^2 \quad \text{--- (III)}$$

From above table using calculators,

$\Sigma x = 720$ $\Sigma x^2 = 47216$

$\Sigma y = 736$ $\Sigma y^2 = 49304$

$\Sigma xy = 48190$ $n = 11$

Substituting these values in (II) & (III) we get

$$736 = 11a + 720b \quad \text{--- (IV)}$$

$$48190 = 720a + 47216b \quad \text{--- (V)}$$

Solving equations (IV) & (V) we get

$$a = 55.51$$

$$\& \ b = 0.174$$

substituting values of ⓐ & ⓑ in (i) we get
$y = 55.51 + 0.174x$ is required equation

c b) The least square regression of line of x on y is given as,
$x = c + dy$ —— (VI)
The normalized equations of (VI) are
$\Sigma x = nc + d\Sigma y$ —— (VII)
& $\Sigma xy = c\Sigma y + d\Sigma y^2$ —— (VIII)
substituting values as calculated in above equations we get
$720 = 11c + 736d$ —— (IX)
$48190 = 736c + 49304d$ —— (X)
solving equations (IX) & (X) we get
$c = 47.90$
& $d = 0.26$
substituting values of c & d in (VI) we get.
$x = 47.90 + 0.26y$ is required equation.

Q) On April in 1994, the following concentration of population were record at eight stations of the monitoring system for air pollution control located in the down area of milan Italy

	stations							
NO_2, mg/m^3	130	130	115	120	135 ~~142~~	142	90	116
CO_2, mg/m^3	2.9	4.4	3.6	4.1	3.3	5.7	4.8	7.3

i) show the relationship between NO_2 & CO_2 by graphical method.
~~sol^n Given,~~

ch. 5

~~let NO_2 be x & CO_2 be y.~~ ii) compute the correlation coefficient between NO_2 & CO_2
iii) Explain relationship between NO_2 & CO_2.

Solⁿ let NO_2 be x & CO_2 be y.

& $y = a + bx$ —— ① be eqⁿ of st. line.

The normalized equations of ① are

$$\Sigma y = na + b\Sigma x \quad \text{—— ②}$$

$$\& \ \Sigma xy = a\Sigma x + b\Sigma x^2 \quad \text{—— ③}$$

From the table & by the use of calculator we get

$\Sigma x = 978$ $\quad \Sigma x^2 = 121370$

$\Sigma y = 36.1$ $\quad \Sigma y^2 = 177.25$

$\Sigma xy = 4388.7$ $\quad n = 8$

substituting above values in ② & ③ we get

$$36.1 = 8a + 978b \quad \text{—— ④}$$

$$\& \ 4388.7 = 978a + 121370b \quad \text{—— ⑤}$$

solving equation ④ & ⑤ we get,

$$a = 6.169$$

$$\& \ b = -0.01355$$

substituting the values of a & b in ① we get

i) ⇒ $y = 6.169 - 0.01355x$

ii) ⇒ correlation coefficient $(r) = \dfrac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y^2)}}$

or, $r = \dfrac{8 \times 4388.7 - 978 \times 36.1}{\sqrt{8 \times 121370 - (978)^2}\sqrt{8 \times 177.25 - (36.1)^2}}$

$\therefore r = -0.1522$

iii) $\Rightarrow$ NO_2 & CO_2 are weakly correlated with each other.

064 shrawan
5) Obtain the equation of the two lines of regression for the following data.

x	43	44	46	40	44	42	45	42	38	40
y	29	31	19	18	19	27	27	24	41	30

Sol^n From the above table using calculator

$\Sigma x = 424$ $\quad \Sigma x^2 = \cancel{18036}$ 18034

$\Sigma y = 265$ $\quad \Sigma y^2 = 7463$

$\Sigma xy = 11156$ $\quad n = 10$

Now, line of regression of y on x is,

$y = a + bx$ —— (I)

The normalized equations of (I) are,

$\Sigma y = na + b\Sigma x$ —— (II)

& $\Sigma xy = a\Sigma x + b\Sigma x^2$ —— (III)

substituting the values of calculated in (II) & (III) we get

$265 = 10a + 424b$ —— (IV)

& $11156 = 424a + 18034b$ —— (V)

Solving equations (IV) & (V) we get

$a = 86.64$

& $b = -1.42$

substituting the values of a & b in (I) we get

$$y = 86.64 - 1.42x$$

Again, line of regression of x on y is

$$x = c + dy \quad \text{—— (VI)}$$

The normalized equations are

$$\Sigma x = nc + d\Sigma y \quad \text{—— (VII)}$$

$$\& \ \Sigma xy = c\Sigma y + d\Sigma y^2 \quad \text{—— (VIII)}$$

Substituting the calculated values in (VII) & (VIII) we get

$$424 = 10c + 265d \quad \text{—— (IX)}$$

$$\& \ 11156 = 265c + 7463d \quad \text{—— (X)}$$

Solving equations (IX) & (X) we get,

$$c = 47.21$$

$$\& \ d = -0.18$$

Substituting values of c & d in (VI) we get

$$x = 47.21 - 0.18y$$

070 Bhadra

Q) A sample of 10 values of three variables x_1, x_2 & x_3 over obtained as,

$\Sigma x_1 = 10$	$\Sigma x_2 = 20$	$\Sigma x_3 = 30$
$\Sigma x_1^2 = 20$	$\Sigma x_2^2 = 68$	$\Sigma x_3^2 = 170$
$\Sigma x_1 x_2 = 10$	$\Sigma x_1 x_3 = 15$	$\Sigma x_2 x_3 = 64$

Find

a) Partial correlation between x_1 & x_3 eliminating the effect of x_2.

b) Multiple correlation between x_2 & x_3 assuming x_1 as independent.

ch. 5

Soln We have,

$$r_{12} = \frac{n\sum x_1x_2 - \sum x_1 \sum x_2}{\sqrt{n\sum x_1^2 - (\sum x_1)^2}\sqrt{n\sum x_2^2 - (\sum x_2)^2}}$$

$$= \frac{10\times 10 - 10\times 20}{\sqrt{10\times 20 - (10)^2}\sqrt{10\times 68 - (20)^2}}$$

$$= -0.5976$$

$$r_{23} = \frac{n\sum x_2x_3 - \sum x_2 \sum x_3}{\sqrt{n\sum x_2^2 - (\sum x_2)^2}\cdot\sqrt{n\sum x_3^2 - (\sum x_3)^2}}$$

$$= \frac{10\times 64 - 20\times 30}{\sqrt{10\times 68 - (20)^2}\cdot\sqrt{10\times 170 - (30)^2}}$$

$$= 0.0845$$

$$r_{13} = \frac{n\sum x_1x_3 - \sum x_1 \sum x_3}{\sqrt{n\sum x_1^2 - (\sum x_1)^2}\times\sqrt{n\sum x_3^2 - (\sum x_3)^2}}$$

$$= \frac{10\times 15 - 10\times 30}{\sqrt{10\times 20 - (10)^2}\sqrt{10\times 170 - (30)^2}}$$

$$= -0.5303$$

Now, i) ⇒ partial correlation between x_1 & x_3 eliminating the effect of x_3 is

$$r_{13.2} \; \cancel{r_{12.3}} = \frac{r_{13} - r_{12}\cdot r_{23}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{23}^2}}$$

$$= \frac{(-0.5303) - (-0.5976)(0.0845)}{\sqrt{1-(-0.5976)^2}\sqrt{1-(0.0845)^2}}$$

$$= -0.60$$

ch. 5

Also, ii) Multiple correlation between x_2 & x_3 assuming x_1 as independent is

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{23}r_{13}}{1-r_{23}^2}}$$

$$= \sqrt{\frac{(-0.5976)^2+(-0.5303)^2-2(-0.5976)(0.0845)(-0.5303)}{1-(0.0845)^2}}$$

$\therefore R_{1.23} = 0.767$

Q) Computer while calculation correlation coefficient between two variable x & y from 25 pair of observation obtained the following results.

$n=25$ $\quad \Sigma x = 125$ $\quad \Sigma x^2 = 650$

$\Sigma y = 100$ $\quad \Sigma y^2 = 460$ $\quad \Sigma xy = 508$

It was however, later discovered at the time of checking that he had copied down two pair as,

x	y
6	14
8	8

while the correct values were

x	y
8	12
6	8

obtain the true value of correlation coefficient.

soln. Here,

corrected $\Sigma x = 125 - 6 - 8 + 8 + 6 = 125$

corrected $\Sigma y = 100 - 14 - 6 + 12 + 8 = 100$

corrected $\Sigma x^2 = 650 - 36 - 64 + 64 - 36 = 650$

corrected $\Sigma y^2 = 460 - 196 - 36 + 144 + 64 = 436$

corrected $\Sigma xy = 508 - 84 - 48 + 96 + 48 = 520$

Now, correlation coefficient $(r) = \dfrac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$

or, $r = \dfrac{25 \times 520 - 125 \times 100}{\sqrt{25 \times 650 - (125)^2}\sqrt{25 \times 436 - (100)^2}}$

$\therefore r = 0.67$

Hence, true value of correlation coefficient is $r = \underline{\underline{0.67}}$.

chapter:-6 (Inference concerning mean). Q) 11 & 12

6.1) ~~Poing~~ Point estimation and interval estimation.

* Estimation.

Estimation is the process by which numerical value is assigned to population parameter based on the information collected from sample.

* Estimator

A sample statistics which is used to estimate a population parameter is called estimator.

i) The sample mean ($\bar{x}$) is an estimator for the population mean (μ).

ii) The sample proportion ($\hat{p}$) is an estimator for the population proportion (P).

iii) The sample standard deviation (s) is an estimator for the population standard deviation (σ).

* Types of estimation

a) Point estimation.

b) Interval estimation.

a) Point estimation.

The process in which a single sample statistic is used to estimate the population parameter is known as point estimation.

Example:- selection of 100 bricks selected randomly from a lot of bricks.

$$\mu = \hat{\mu} = \mu_{\bar{x}} = \bar{x} = \frac{\Sigma x_i}{n}$$

$$\sigma^2 = E(s^2) = \frac{n\Sigma x_i^2 - (\Sigma x_i)^2}{n(n-1)}$$

b) Interval estimation (I.E.)

As point estimates cannot really be expected to coincide with the quantities intended to estimate, it is sometime needed to replace them with interval estimates that is with intervals for which we can ~~start~~ assert with a reasonable degree of certainty that they will contain the parameter under consideration such types of estimation is called interval estimation.

I.E = point estimation $\pm$ (standard error of the mean) $\times Z_{\alpha/2}$

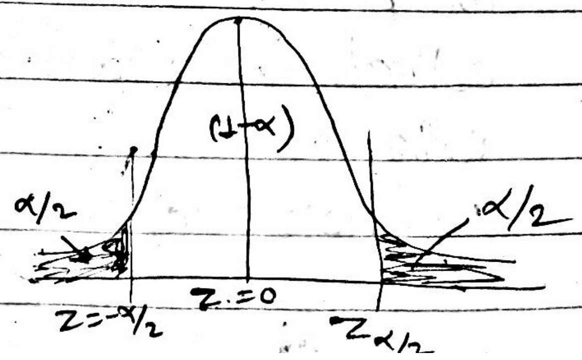
(If the sample are from normal population)

$$= \bar{x} \pm \frac{\sigma}{\sqrt{n}} Z_{\alpha/2} \text{ for } (100-\alpha)\% \text{ confidence level of } \mu.$$

$$= \bar{p} \pm \sqrt{\frac{\bar{p}\bar{q}}{n}}\, Z_{\alpha/2}$$

$$= \bar{x} \pm \frac{s}{\sqrt{n}} Z_{\alpha/2} \quad \text{for } n \geq 30$$

$$= \bar{x} \pm \frac{s}{\sqrt{n}} t_{\alpha/2, n-1} \quad \text{for } n < 30$$



chaitra

Q) An analysis for pH (acidity) in an random sample of water from 40 rainfalls showed that mean is 6.7 and s.d. is 0.5. Find a 99% confidence interval for the mean pH in rainfalls.

soln Given, $n = 40$, $\bar{x} = 6.7$, $s = 0.5$

Now, for 99% confidence level, $\alpha = 1\%$.

i.e $\alpha = 0.01$ then $z_{\alpha/2} = z_{0.005} = 2.576$

The required ~~regression~~ confidence level for population μ is

$$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} z_{\alpha/2}$$

$$= 6.7 \pm \frac{0.5}{\sqrt{40}} \times (2.576)$$

$$= 6.7 \pm 0.20365$$

$$= (6.90365, 6.49635)$$

i.e $6.5 \le \mu \le 6.90$

Q) A sample of 900 members has a mean 3.5 cm & s.d 2.61. If the population in normal & its mean is unknown find the 95% and 98% fiducial limits of true mean.

soln Given, $n = 900$, $\bar{x} = 3.5$, $s = 2.61$

Now, for 95% fiducial limits of true mean, $\alpha = 5\%$

i.e $\alpha = 0.05$ then $z_{\alpha/2} = z_{0.025} = 1.960$

The required ~~regression~~ confidence level for population μ is

$$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} z_{\alpha/2}$$

or, $\mu = 3.5 \pm \frac{2.61}{\sqrt{900}} \times (1.960)$

or, $\mu = 3.5 \pm 0.17052$

or, $\mu = (3.67052, 3.32948)$

i.e $3.33 \le \mu \le 3.67$

Also,

for 98% fiducal limit for the population mean, $\alpha = 2\%$

i.e $\alpha = 0.02$ then $Z_{\alpha/2} = \cancel{Z_{0.005}} = Z_{0.01} = 2.326$

The required confidence level for population μ is

$$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} Z_{\alpha/2}$$

or, $\mu = 3.5 \pm \frac{2.61}{\sqrt{900}} \times 2.326$

or, $\mu = 3.5 \pm 0.202362$

or, $\mu = (3.702362, 3.297638)$

i.e $3.3 \le \mu \le 3.7$

071 shravan.

8) The mean weight loss of $n = 16$ grinding balls after a certain length of time in mill slurry is 3.42 grams with a standard deviation of 0.68 gram. Construct a 99% confidence interval for the true mean weight loss of such grinding balls under the stated conditions.

Soln Given, $n = 16$, $\bar{x} = 3.42$ grams, $s = 0.68$ grams

Now, for 99% confidence level i.e $\alpha = 1\% = 0.01$

$$Z_{\alpha/2} = Z_{0.005} = 2.576$$

∴ The required confidence interval is given as,

$$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} Z_{\alpha/2}$$

or, $\mu = 3.42 \pm \frac{0.68}{\sqrt{16}} \times 2.576$

or, $\mu = 3.42 \pm 0.43792$

or, $\mu = (3.85792, 2.98208)$

i.e $2.98 \leq \mu \leq 3.86$.

069 chaitra

Q) A random sample of size 16 showed a mean of 52 with a standard deviation 4. Obtain 99% & 95% confidence limits population mean.

Sol^n Given, $n = 16$, $\bar{x} = 52$, $s = 4$

Now, for 99% confidence limit

i.e $\alpha = 1\% = 0.01$ so, $Z_{\alpha/2} = Z_{0.005} = 2.576$

So The required confidence limits population mean

$$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} Z_{\alpha/2}$$

or, $\mu = 52 \pm \frac{4}{\sqrt{16}} \times 2.576$

or, $\mu = 52 \pm 2.576$

or, $\mu = (54.576, \cancel{50.576}\ 49.424)$

i.e ~~$50.576 \leq \mu \leq 54.576$~~ 49.424.

i.e $49.424 \leq \mu \leq 54.576$ //

Also, For 95% confidence limit

i.e $\alpha = 5\% = 0.05$ $\therefore Z_{\alpha/2} = Z_{0.025} = 1.960$

$\therefore$ The required confidence limit population mean is,

$$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} Z_{\alpha/2}$$

or, $\mu = 52 \pm \frac{4}{\sqrt{16}} \times 1.96$

or, $\mu = 52 \pm 1.96$

or, $\mu = (53.96, 50.04)$

i.e $50.04 \le \mu \le 53.96.$

068 chaitra

Q) A sample of 900 members has a mean of 3.4 cm & standard deviation of 2.61 cm. If the population is normal and its mean is unknown, find 95% & 98% fidcial limits of the true mean.

Soln Given, $n = 900$, $\bar{x} = 3.4$ cm, $s = 2.61$ cm

Now, for 95% fidcial limits for the population mean μ.

i.e $\alpha = 5\% = 0.05$, $Z_{\alpha/2} = Z_{0.025} = 1.96$

$\therefore$ The confidence interval for the population mean is

$$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} Z_{\alpha/2}$$

or, $\mu = 3.4 \pm \frac{2.61}{\sqrt{900}} \times 1.96$

or, $\mu = 3.4 \pm 0.17052$

or, $\mu = (\cancel{3.22948}, 3.57052, \cancel{0.22}\ 3.22948)$

i.e $3.23 \le \mu \le 3.57$

Also,

for 98% fidcial limits for the population mean μ.

i.e $\alpha = 2\% = 0.02$ $\therefore Z_{\alpha/2} = Z_{0.01} = 2.326$ ~~(2.326)~~

$\therefore$ The confidence interval for the population mean is.

$$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} Z_{\alpha/2}$$

$$\text{or, } \mu = 3.4 \pm \frac{2.61}{\sqrt{900}} \times 2.326$$

or, $\mu = 3.4 \pm 0.202362$

or, $\mu = (3.602362, 3.197638)$

i.e $3.20 \le \mu \le 3.60$

6.2) Test of hypothesis.

Hypothesis is tentative assumption about population parameter. Testing of hypothesis is statistical procedure that involved of formulating the hypothesis and testing of validity of hypothesis.

Hypothesis are the assumptions or guess about the populations involved. Such assumptions may or maynot be true.

There are two types of hypothesis, they are

i) Null Hypothesis.

ii) Alternative hypothesis.

a) Null hypothesis

Null hypothesis is initially assumed to be true, which indicates that there is no significance difference between the sample statistic and population parameter. It is denoted by H_o.

Here, $H_o: \theta = \theta_o$

where, θ_o = specified value of population parameter which shows that null hypothesis is expressed as an equality.

b) Alternate hypothesis

It is a hypothesis which is accepted if the null hypothesis is rejected. it is complementary hypothesis of null hypothesis. Alternate hypothesis is denoted by H_1 & defined as,

$H_1: \theta \neq \theta_o$

which shows that there is significance difference between the sample statistic & population parameter.

This type of test is called two tailed test.

If $H_1: \theta > \theta_o$, then population parameter is greater than specified value, also called Right tailed test or one tailed test.

If $H_1: \theta < \theta_o$, then population parameter is less than specific value, also called left tailed test or one tailed test

This can be summarized as,

$H_1: \theta \neq \theta_0$ (two tailed test)

$H_1: \theta > \theta_0$ (right tailed test)

$H_1: \theta < \theta_0$ (left tailed test)

} (one tailed test).

* Types of Errors in hypothesis.
  i) Type - I error
  ii) Type - II error

i) Type - I error

If we rejects the Null hypothesis H_0 when it is true, the error occured from this decision is type-I error.

The probablity of making type-I error is denoted by α (level of significance).

i.e P(Reject H_0 when H_0 is true) $= P$(Type-I error) $= \alpha$

where α is called size of error of type-I error.

whereas, the complement of α i.e $(1-\alpha)$ is called the confidence coefficient. Then $(1-\alpha)$ is the probablity of accepting H_0 when H_0 is true.

ii) Type - II error

This type of error occured while accepting Null hypothesis which is false. The probablity of making type-II error is β.

i.e. P(Accept H_0 when it is false) $= P$(type II error) $= \beta$.

where β is called size of error of type-II error.

* level of significance

The maximum size of type-I error which are prepared to risk is known as level of significance mathematically,

$$\alpha = P(\text{Type-I error})$$

Usually 1% & 5% significance level is used. The level of significance 5% indicates that we are ready to take 5% risk of rejecting true Null hypothesis, H_0. i.e probability of rejecting true $H_0 = 0.05$

* Working rule to test a hypothesis.

The summary of the hypothesis test is given below.

1) The parameter of interest.
2) Two types of hypothesis.
3) Level of significance (α).
4) The test statistics = $\dfrac{\text{statistics} - \text{parameter}}{\text{standard error of statistics}}$

i) $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ (normal population σ known)

ii) $z = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ (normal population σ unknown, $n \geq 30$)

iii) $z = \dfrac{\bar{p} - P_0}{\sqrt{\dfrac{P_0 q_0}{n}}}$ (population proportion with $nP_0 \geq 5$, $nq_0 \geq 5$)

iv) $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ (normal population, $n < 30$, σ unknown)

5) Acceptance / Rejection region.

Alternate hypothesis	critical region for $\mu = \mu_0$ & $n \geq 30$ Rejected H_0 if	critical region for $\mu = \mu_0$ & $n < 30$ Rejected H_0 if
left tail test, $\mu < \mu_0$, $P < P_0$	$z \leq Z_{-\alpha}$	$t \leq -t_{\alpha, n-1}$
right tail test $\mu > \mu_0$, $P > P_0$	$z \geq Z_{\alpha}$	$t \geq t_{\alpha, n-1}$
two tail test $\mu \neq \mu_0$, $P \neq P_0$	$\lvert z \rvert \geq Z_{\alpha/2}$	$\lvert t \rvert \geq t_{\alpha/2, n-1}$

6) Decision.

6.3) Hypothesis test concerning one mean.

A) Hypothesis test for large sample ($n \geq 30$) [z-test].

Procedure for testing hypothesis in this condition is.

1) The parameter of interest

2) set up two types of hypothesis

Null hypothesis, $H_0 : \mu = \mu_0$

Alternative hypothesis, $H_1 : \mu \neq \mu_0$ (two tailed test).

$H_1 : \mu > \mu_0$ (Right tailed test).

$H_1 : \mu < \mu_0$ (Left tailed test).

3) Level of significance (α).

choose an appropriate level of significance if α is not specified we usually use $\alpha = 5\%$.

4) Test statistics.

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \quad \text{if } \sigma^2 \text{ is known}$$

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad \text{if } \sigma^2 \text{ is unknown}$$

where $s^2 = 1/n \, \Sigma (x_i - \bar{x})^2$

5) criteria

obtain critical value from table (i.e z_α as $z_{\alpha/2}$) at pre-specified level of significance.

6) Decision.

i) if $|z| > z_{\alpha/2}$ (for two tailed test)

if $|z| > z_\alpha$ (for one tailed test)

it is significant and reject H_0 & hence accept H_1.

ii) if $|z| < z_{\alpha/2}$ (for two tailed test)

if $|z| < z_\alpha$ (for one tailed test)

it is not significant & accept H_0 & hence reject H_1.

Note:-

i) If $(\bar{x} - \mu_0) = 0$, & if our test suggests rejection of the null hypothesis we commit type-I error.

ii) If $(\bar{x} - \mu_0) = 0$ & if our test suggests acceptance of the null hypothesis we commit type II error.

B) Hypothesis test for small ($n < 30$) [t-test].

Procedure for testing hypothesis under this condition is

1) The parameter of interest
2) set up two types of hypothesis.

Null hypothesis, $H_0 : \mu = \mu_0$

Alternative hypothesis, $H_1 : \mu \neq \mu_0$ (two tailed test)

$H_1 : \mu > \mu_0$ (Right tailed test)

$H_1 : \mu < \mu_0$ (Left tailed test)

3) Level of significance (α).
4) Test statistics,

$$t = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n-1}} \quad \text{if } \sigma^2 \text{ is given}$$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad \text{if } \sigma^2 \text{ is not given}$$

where $s^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2$

5) critical value.

obtain critical value from table of t-distribution at level of significance (α) & degree of freedom ($n-1$) according to one-tailed or two tailed (i.e. $t_{\alpha, n-1}$ or $t_{\alpha/2, n-1}$).

6) Decision

i) If $|t| \geq t_{\alpha/2, n-1}$ (for two tailed test)

if $|t| > t_{\alpha, n-1}$ (for one tailed test)

it is significant & ~~accept~~ reject H_0 & reject H_1.

ii) If $|t| < t_{\alpha/2, n-1}$ (for two tailed test)

if $|t| \leq t_{\alpha, n-1}$ (for one tailed test)

it is not significant & ~~accept~~ reject H_1 & reject H_0.

6.4) Hypothesis test concerning two mean.

A) Hypothesis test for large sample ($n \geq 30$).

Procedure for testing hypothesis under this condition is

~~1) The parameter of interest.~~

1) set up Null hypothesis

$H_0: \mu_1 = \mu_2$

2) set up Alternate hypothesis

$H_1: \mu_1 \neq \mu_2$ (two tailed test)

or, $H_1: \mu_1 > \mu_2$ (right tailed test)

or, $H_1: \mu_1 < \mu_2$ (left tailed test).

3) level of significance (α).

choose the appropriate level of significance or take $\alpha = 5\%$ if not specified.

4) Test statistic

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$ if σ_1^2 & σ_2^2 are known

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ if σ_1^2 & σ_2^2 are not known

where $s_1^2 = 1/n_1 \, \Sigma (X_{1i} - \bar{X}_1)^2$

$s_2^2 = 1/n_2 \, \Sigma (X_{2i} - \bar{X})^2$

5) critical value ($Z_{\alpha/2}$ or Z_α)

obtain critical values (i.e. $Z_{\alpha/2}$ or Z_α) from table of z-distribution at specified level of significance.

6) Decision

i) if $|z| > z_{\alpha/2}$ (for two tailed test)
if $|z| > z_{\alpha}$ (for one tailed test)
z-lies in rejection region so we reject H_0 & accept H_1.

ii) if $|z| < z_{\alpha/2}$ (for two tailed test)
if $|z| < z_{\alpha}$ (for one tailed test)
z-lies in acceptance region so we accept H_0 & reject H_1.

Note:-

- If first sample statistic < second sample statistic, then we use left tailed test
- If first sample statistic > second sample statistic, then we use right tailed test.

B) Hypothesis test for small sample ($n < 30$).
procedure of testing hypothesis under this condition is

1) set-up Null hypothesis.
$H_0 : \mu_1 = \mu_2$

2) set up Alternate hypothesis
$H_1 : \mu_1 \neq \mu_2$ (two tailed test)
$H_1 : \mu_1 > \mu_2$ (right tailed test)
$H_1 : \mu_1 < \mu_2$ (left tailed test).

3) level of significance (α)
take $\alpha = 5\%$ if not given

4) Test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ if σ_1^2 & σ_2^2 are not known & are not equal

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ if σ_1^2 & σ_2^2 are not known but are equal.

where,

$$s_1^2 = \frac{1}{n_1 - 1} \Sigma (x_{1i} - \bar{x})^2$$

$$s_2^2 = \frac{1}{n_2 - 2} \Sigma (x_{2i} - \bar{x})^2$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

5) Critical value ($t_{\alpha/2}$ or t_α)

obtain critical value from t-distribution table at specific level of significance & degree of freedom (v)

$v = n_1 + n_2 - 2$ if σ_1^2 & σ_2^2 are equals

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n+1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

6) Decision

i) if $|t| > t_{\alpha/2}$ (for two tailed test)

if $|t| > t_\alpha$ (for one tailed test)

t lies in rejection region so we reject ~~Null~~ H_0 & accept H_1

ii) if $|t| \leq t_{\alpha/2}$ (for two tailed test)

if $|t| \leq t_\alpha$ (for one tailed test)

t lies in acceptance region so we accept H_0 & reject H_1

* Paired t-test (before & after aayo vane).

4) Test statistics $t = \frac{\bar{D}}{S_D/\sqrt{n}}$;

$$\bar{D} = \frac{\Sigma D_i}{n} \quad ; \quad D_i = X - Y$$

$$S_D = \sqrt{\frac{n\Sigma D_i^2 - (\Sigma D_i)^2}{n(n-1)}} \qquad t_{\alpha, n-1}$$

071 chaitra.

Q) In a certain factory, there are two independent processes manufacturing the same item. The average weight in a sample of 250 items produced from one process is found to be 120 gram with a standard deviation of 12 gram, while the corresponding figures in a sample of 400 items from the other process are 124 & 14 respectively. Test whether the two mean weights differ significantly or not at 5% level of significance?

Soln Procedure of testing hypothesis is given as,

1) Null hypothesis:-
$H_0: \mu_1 = \mu_2$ difference of mean is not significant

2) Alternate hypothesis:-
$H_0: \mu_1 \neq \mu_2$

3) level of significance (α)
$\alpha = 5\% = 0.05$

4) Test statistic

8) Define confidence level & significance level.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ for $n_1 \geq 30$ & $n_2 \geq 30$

5) Criteria

we reject H_0 if $|z| \geq 1.960$

6) Calculation & decision

Given,

$n_1 = 250$, $\bar{x}_1 = 120$, $s_1 = 12$

$n_2 = 400$, $\bar{x}_2 = 124$, $s_2 = 14$

so,

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{120 - \cancel{400}124}{\sqrt{\frac{(12)^2}{250} + \frac{(14)^2}{400}}} = -3.87$$

Since, $|z| = 3.87$ is greater than $z_{0.025} = 1.96$, therefore we reject H_0 & accept H_1. We conclude that difference of mean is significant.

---

071 chaitra

8) Define confidence level & significance level. A company claims that its light bulbs are superior to those of its main competitor. If a study showed that a sample of 40 of its bulbs has mean lifetime of 647 hours of continuous use with standard deviation of 27 hours. While a sample of 40 bulbs made by its main competitor had mean lifetime of 638 hours of continuous use with standard deviation of 31 hours. Does this substantiate

claim at 1% level of significance?

Soln. Procedure of testing hypothesis is given as.

1) Null hypothesis

$H_0: \mu_1 = \mu_2$ difference of mean is not significant.

2) Alternate hypothesis

$H_1: \mu_1 > \mu_2$ (one tailed test)

3) level of significance (α)

$\alpha = 1\% = 0.01$

4) Test statistics

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ for $n_1 \geq 30$ & $n_2 \geq 30$

5) criteria

we reject H_0 if $|z| \geq 2.326$

6) calculation & decision

Given,

$n_1 = 40$, $\bar{x}_1 = 647$, $s_1 = 27$

$n_2 = 40$, $\bar{x}_2 = 638$, $s_2 = 31$

So,

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{647 - 638}{\sqrt{\frac{(27)^2}{40} + \frac{(31)^2}{40}}} = 1.3846$$

since $|z| = 1.3846$ is less than $z_{0.01} = 2.326$, therefore we accept H_0. Hence we conclude that both company are significant.

~~his claim is not substantiate~~

069 chaitra

Q) In a manufacturing company the new modern manager is in a belief that music enhances the productivity of workers. He made observations on 6 workers for a week and recorded the production before and after the music was installed. From the data given below, can you conclude that the productivity has indeed changed due to music? ($\alpha = 1\%$).

week without music	219	205	226	198	209	216
week with music	235	186	240	203	221	205

Soln Let x_i and y_i, $i = 1, 2, \ldots, n$ be the value of production before and after the music was installed.

Null hypothesis:-

$H_0 : \mu_D = 0$

Alternative hypothesis:-

$H_1 : \mu_D > 0$

Level of significance

$\alpha = 1\% = 0.01$

test statistic

$$t = \frac{\bar{D}}{S_D/\sqrt{n}}$$

For $\bar{D}$ & S_D

x_i Before	219	205	226	198	209	216	
y_i After	235	186	240	203	221	205	
$D_i = x_i - y_i$	-16	19	-14	-5	-12	11	$\Sigma D_i = -17$
D_i^2	256	361	196	25	144	121	$\Sigma D_i^2 = 1103$

so, $\bar{D} = \frac{\sum D_i}{n} = \frac{-17}{6} = -2.833$

$$S_D = \sqrt{\frac{n\sum D_i^2 - (\sum D_i)^2}{n(n-1)}}$$

$$= \sqrt{\frac{6 \times 1103 - (-17)^2}{6(6-1)}}$$

$$= 14.52$$

$$\therefore t = \frac{\bar{D}}{S_D/\sqrt{n}} = \frac{-2.833}{14.52} \times \sqrt{6} = -0.4779 \approx -0.5$$

criteria

we reject H_0 if $|t| > t_{\alpha, n-1}$ i.e $|t| > t_{0.01,5}$

$|t| > 3.365$

conclusion

calculated value of $|t|$ is less than 3.365. Therefore we ~~accept~~ ~~reject~~ accept null hypothesis H_0. We may conclude that ~~students~~ music have not enhance the productivity of workers.

071 shrawan

Q) The following are the average weekly losses of workers hours due to accidents in 10 industrial plants before and after a certain safety program was put into operation.

Before	45	73	46	124	33	57	83	34	26	17
After	36	60	44	119	35	51	77	29	24	18

use the 0.05 level of significance to test whether the safety program is effective.

sol^n let x_i and y_i, $i = 1, 2, 3, \ldots n$ be the value of average hour losses before & after the safety program was put into operation.

Null hypothesis

$H_0: \mu_D = 0$ i.e safety program donot get effective

Alternate hypothesis

$H_1: \mu_D > 0$

level of significance

$\alpha = 0.05$

test statistics

$$t = \frac{\bar{D}}{S_D/\sqrt{n}}$$

For $\bar{D}$ & S_D

x_i Before	45	73	46	124	33	57	83	34	26	17
y_i After	36	60	44	119	35	51	77	29	24	11
$D_i = x_i - y_i$	9	13	2	5	-2	6	6	5	2	6
D_i^2	81	169	4	25	4	36	36	25	4	36

$\Sigma D_i = 52$

$\Sigma D_i^2 = 420$

so, $\bar{D} = \frac{\Sigma D_i}{n} = \frac{52}{10} = 5.2$

$$\& \; S_D = \sqrt{\frac{n\Sigma D_i^2 - (\Sigma D_i)^2}{n(n-1)}}$$

$$= \sqrt{\frac{10 \times 420 - (52)^2}{10(10-1)}}$$

$$= 4.077$$

$$\therefore t = \frac{\bar{D}}{S_D/\sqrt{n}} = \frac{5.2}{4.077} \times \sqrt{10} = 4.033$$

criteria

we reject H_0 if $|t| > t_{\alpha, n-1}$ i.e $|t| > t_{0.05, 9}$

i.e $|t| > 1.833$

conclusion

calculated value of $|t|$ is greater than 1.833. Therefore we ~~reject~~ accept reject H_0 ~~& accept H_1~~. We may conclude that the safety program was ~~effective.~~ ~~not~~ effective.

Q) Eleven college students were given a test in statistics. They were given a month's tuition & a second test was held at the end of it. Do the marks give evidence that the students have benefited by the extra coaching? (use $\alpha = 0.05$).

marks in 1st test	23	20	19	21	18	20	18	17	23	16	19	24
marks in 2nd test	24	19	22	18	20	22	20	20	23	20	18	22

Soln let x_i & y_i, $i = 1, 2, 3, \dots n$ be the values of marks obtained by students before & after the tuition.

Null hypothesis:-

$H_0: \mu_D = 0$ i.e there is no benefited by tuition.

Alternative hypothesis:-

$H_1: \mu_D > 0$

level of significance

$\alpha = 0.05$

test statistics

$$t = \frac{\bar{D}}{S_D/\sqrt{n}}$$

for $\bar{D}$ & S_D

x_i: Before	23	20	19	21	18	20	18	17	23	16	19	24
y_i: After	24	19	22	18	20	22	20	20	23	20	18	22
$D_i = x_i - y_i$	-1	1	-3	3	-2	-2	-2	-3	0	-4	1	2
D_i^2	1	1	9	9	4	4	4	9	0	16	1	4

$\Sigma D_i = -10$

$\Sigma D_i^2 = 62$

so, $\bar{D} = \frac{\Sigma D_i}{n} = \frac{-10}{12} = -0.833$

& $S_D = \sqrt{\frac{n\Sigma D_i^2 - (\Sigma D_i)^2}{n(n-1)}}$

$= \sqrt{\frac{12 \times 62 - (-10)^2}{12(12-1)}}$

$= 2.2088$

$\therefore t = \frac{\bar{D}}{S_D/\sqrt{n}} = \frac{-0.833}{2.2088} \times \sqrt{12} = -1.3064$

criteria

we reject H_0 if $|t| > t_{\alpha, n-1}$ i.e $|t| > t_{0.05, 11}$

i.e $|t| > 1.796$

conclusion

calculated value of $|t|$ is less than 1.796, Therefore we accept H_0. We may conclude that there is no any benefited by extra coaching.

068 chaitra

5) A potential buyer of light bulbs bought 50 bulbs each of two brands. Upon testing these bulbs, he found that brand A had a mean life of 1282 hours with S.D. of 80 hours whereas the brand B had a mean life of 1208 hours with S.D. of 94 hours. can the buyer be quite certain that the two brands do differ in quality? $\alpha = 10\%$.

Sol^n Procedure of testing hypothesis is given as,

Null hypothesis :-

$H_0 : \mu_1 = \mu_2$ difference of mean is not significant

Alternate hypothesis :-

$H_1 : \mu_1 \neq \mu_2$ (two tailed test)

level of significance

$\alpha = 10\%$ i.e $\alpha = 0.10$

Test statistics

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad \text{for } n_1 \geq 30 \ \& \ n_2 \geq 30$$

criteria

we reject H_0 if $|z| > 1.645$

calculation & decision

Given,

$n_1 = 50$, $\bar{X}_1 = 1282$, $S_1 = 80$

$n_2 = 50$, $\bar{X}_2 = 1208$, $S_2 = 94$

$$\text{so, } Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{1282 - 1208}{\sqrt{\frac{(80)^2}{50} + \frac{(94)^2}{50}}} = 4.239$$

since $|z|$ is greater than $z_{0.05} = 1.645$ so we reject H_0.
Hence we conclude that two brands is differ in quality

Q) A random sample of 100 recorded death in a certain hospital during the past years showed an average life span of 71.8 years, with a standard deviation of 8.9 years. Does this seems to indicate that the average life span today is less than 75 years? take $\alpha = 0.05$.

Soln Given, size of random sample, $n = 100$
sample mean, $\bar{x} = 71.8$ years
sample standard deviation, $s = 8.9$ years
specified mean, $\mu_0 = 75$ years
level of significance $= 0.05$

Now,
Null hypothesis:
$H_0: \mu = \mu_0 = 75$, ~~the population mean~~ difference of mean is not significant

Alternate hypothesis:-
$H_1: \mu < 75$ (one tailed test)

level of significance
$\alpha = 0.05$

test statistics
for $n \geq 30$
$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad \cancel{\frac{71.8 - 75 \times \sqrt{100}}{8.9}} \cancel{= 3.59}$$

~~since, |z| is~~ criteria

we reject H_0 if $|z| > 1.645$

calculation & decision

~~Given~~ $z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{71.8 - 75}{8.9} \times \sqrt{100} = -3.59.$

since $|z|$ is greater than 1.645 so, we reject H_0 & accept H_1. We conclude that difference of mean is significant.

072 chaitra.

Q) Define the central limit theorem. A sample of 100 mobile battery cells tested to find the length of life produced the following results as mean 13 months & standard deviation of 3 months. Assuming the data to be normally distributed by using central limit theorem, what percentage of battery cells expected to have Average life? i) More than 15 months
ii) less than 9 months.

Soln If $\bar{x}$ is the mean of a sample of size 'n' taken from a population having the mean 'μ' & finite variance σ^2, then ~~standard normal distribution as $n\to\infty$~~.

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

is a random variable distribution function approches that of the standard normal distribution for $n\to\infty$ $(n>30)$.

Numerical Part

Given, mean $(\mu) = 13$.

standard deviation $(\sigma) = 3$, $n = 100$

i) $P(\bar{x} \geq 15) = ?$

ii) $P(\bar{x} \leq 9) = ?$

Now, $\dfrac{\sigma}{\sqrt{n}} = \dfrac{3}{\sqrt{100}} = \dfrac{3}{10} = 0.3$

So, i) $P(\bar{x} \geq 15) = P(a \leq \bar{x})$

$= 1 - P(\bar{x} \leq a)$

$= 1 - \phi\left(\dfrac{a-\mu}{\sigma/\sqrt{n}}\right)$

$= 1 - \phi\left(\dfrac{15-13}{0.3}\right)$

$$= 1 - \phi(6.67)$$
$$=$$

ii) $P(\bar{X} \leq 9) = P(\bar{X} \leq b)$

$$= \phi\left(\frac{b-\mu}{\sigma/\sqrt{n}}\right)$$

$$= \phi\left(\frac{9-13}{0.3}\right)$$

$$= \phi(-13.33)$$

$$=$$

071 chaitra

7. Q) State central limit theorem. An electrical firm manufactures light bulbs that have a length of life that is approximately normal distribution with mean equal to 800 hours & standard deviation of 4 hours. Find the probablity that a random sample of 16 bulbs will have an average life of less than 12775 hours?

Sol^n ⇒ If $\bar{x}$ is the mean of a sample of size 'n' taken from a population having the mean 'μ' & finite variance σ^2, then according to central limit theorem,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{N}}$$

is a random variable distribution function approches that of the standard normal distribution, for $n \to \infty$ or, $(n > 30)$.

soln Given, mean(μ) = 800

standard deviation (σ) = 4

sample number (n) = 16

$P(\bar{x} \leq 12775) = ?$

Now,

since $n \leq 30$ so central limit theorem cannot be used

so, $P(\bar{x} \leq 12775) = P(\bar{x} < b)$

$$= \phi\left(\frac{b-\mu}{\sigma}\right)$$

$$= \phi\left(\frac{12775-800}{4}\right)$$

$$= \phi(2993.75)$$

$$=$$

Q) State central limit theorem. A random sample of size 100 is taken from an infinite population having the mean 76 and variance 256. What is the probablity that the sample mean will be between 75 & 78?

⇒ If $\bar{x}$ is the mean of a sample of size 'n' taken from a population having the mean 'μ' and finite variance σ^2, then according to central limit theorem,

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{N}}$$

is a random variable distribution function approches that of the standard normal distribution, for $n \to \infty$ or $(n > 30)$.

Numerical Part

Sol^n Given, sample size $(n) = 100$

sample mean $(\bar{x}) = 76$

sample variance $(\sigma^2) = 256$

$\therefore$ sample standard deviation $(\sigma) = \sqrt{256} = 16$

Now, $\frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{100}} = \frac{16}{10} = 1.6$

So, $P(75 \leq \bar{x} \leq 78) = P(a \leq \bar{x} \leq b)$

$$= \phi\left(\frac{b-\mu}{\sigma/\sqrt{n}}\right) - \phi\left(\frac{a-\mu}{\sigma/\sqrt{n}}\right)$$

$$= \phi\left(\frac{78-76}{1.6}\right) - \phi\left(\frac{75-76}{1.6}\right)$$

$$= \phi(1.25) - \phi(-0.6)$$

$$= 0.8944 - 0.2743$$

$$= 0.6201$$

Hence, the probablity that the sample mean will be between 75 & 78 is 0.6201.

ch. 1 Baye's theorem

070 chaitra.

Q) State Baye's theorem. A manufacture of air-conditioning units purchases 70% of its thermostats from company A, 20% from company B and the rest from company C. Past experience shows that 0.5% of company A's thermostats, 1% of company B's thermostats & 1.5% of company C's thermostats are likely to be defective. An air-conditioning unit randomly selected from this manufacture's production line was found to have a defective thermostat. Find the probablity that company A supplied the defective thermostat.

⇒ let $B_1, B_2, B_3, \dots B_n$ be a mutually disjoint events of sample space S & A be any event that occurs with $B_1, B_2, B_3, \dots, B_n$ then Baye's theorem states that

$$P(B_r/A) = \frac{P(B_r \cap A)}{P(A)} = \frac{P(B_r) \cdot P(A/B_r)}{P(A)}$$

$$= \frac{P(B_r) \cdot P(A/B_r)}{\sum_{i=1}^{n} P(B_i) \cdot P(A/B_i)}$$

where, $r = 1, 2, 3, 4, \dots n$.

Numerical Part

soln Given,

$P(A) = 70\% = 0.70$

$P(B) = 20\% = 0.20$

$P(C) = 10\% = 0.10$

if x is the event that product is defective then

$P(x/A) = 0.5 = 0.005$

$P(x/B) = 1\% = 0.01$

Q) Distinguish between mutually exclusive & equally likely events with examples.

Ch.1 Baye's theorem.

$P(x/c) = 1.5\% = 0.015$

$P(A/x) = ?$

Now,

we have

$P(x) = P(A \cap x) + P(B \cap x) + P(C \cap x)$

$= P(A) \cdot P(x/A) + P(B) \cdot P(x/B) + P(C) \cdot P(x/c)$

$= 0.70 \times 0.005 + 0.20 \times 0.01 + 0.10 \times 0.015$

$= 0.007$

so, i) $\Rightarrow P(A/x) = \dfrac{P(A \cap x)}{P(x)}$

$= \dfrac{P(A) \cdot P(x/A)}{P(x)}$

$= \dfrac{0.70 \times 0.005}{0.007}$

$\therefore P(A/x) = 0.5.$

Hence, the probablity that company A supplied the defective thermostat is 0.50.

070 Ashad

Q) Distinguish between mutually exclusive & equally likely events with examples. What is the use of Baye's theorem of probablity? In a college 45% students belongs to civil, 30% Electronics & remaining to other faculties. The probablity of being top is 5%, 4% & 2% respectively in civil, electronics & others. If the year's result is published, what is the probablity that the topper is

8) what is the use of Baye's theorem in theory of probablity?

from electronics?

Sol^n Given,

$P(C) = 45\% = 0.45$, civil students

$P(E) = 30\% = 0.30$, electronics students

$P(O) = 25\% = 0.25$, other students

if x is the event that student is topper, then

$P(x/C) = 5\% = 0.05$

$P(x/E) = 4\% = 0.04$

$P(x/O) = 2\% = 0.02$

& $P(E/x) = ?$

Now, we have

$$p(x) = P(C \cap x) + P(E \cap x) + P(O \cap x)$$
$$= P(C) \cdot P(x/C) + P(E) \cdot P(x/E) + P(O) \cdot P(x/O)$$
$$= 0.45 \times 0.05 + 0.30 \times 0.04 + 0.25 \times 0.02$$
$$= 0.0395$$

so, $$P(E/x) = \frac{P(E \cap x)}{P(x)}$$
$$= \frac{P(E) \cdot P(x/E)}{P(x)}$$
$$= \frac{0.30 \times 0.04}{0.0395}$$

$\therefore P(E/x) = 0.3038$

Hence, the probablity that the topper is from electronics is 0.3038.

Q) Define dependent & independent events with examples

ch. 1 Baye's theorem

o8 chaitra

Q) Define dependent & independent events with examples. In a bolt factory, machines A, B and C manufacture 25%, 35% & 40% of the total respectively. Of their output 5, 4 & 2 percent are defective bolts. A bolt is drawn at random from the product & is found to be defective. What is the probability that it was manufactured from the machine B?

Sol^n Given, $P(A) = 25\% = 0.25$

$P(B) = 35\% = 0.35$

$P(C) = 40\% = 0.40$

if x is the event that produce defective bolts then

$P(x/A) = 5\% = 0.05$

$P(x/B) = 4\% = 0.04$

$P(x/C) = 2\% = 0.02$

$P(B/x) = ?$

Now, we have,

$$P(x) = P(A \cap x) + P(B \cap x) + P(C \cap x)$$

$$= P(A) \cdot P(x/A) + P(B) \cdot P(x/B) + P(C) \cdot P(x/C)$$

$$= 0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02$$

$$= 0.0345$$

So, $$P(B/x) = \frac{P(B \cap x)}{P(x)}$$

$$= \frac{P(B) \cdot P(x/B)}{P(x)}$$

$$= \frac{0.35 \times 0.04}{0.0345}$$

$\therefore P(B/x) = 0.4058$

Hence, the probablity that the bolt was manufactured by the machine B is 0.4058.

058, 061, 068 Bhadra,

8) There are three machines A, B & C producing 1000, 2000 & 3000 articles per hour respectively. These machines are known to be producing 1%, 2% & 3% defectives respectively. One article is selected at random from an hour production of the three machines & found to be defective. What is the probablity that the article is produced from a) Machine A
b) machine B
c) machine C.

Soln Given, $P(A) = 1000/6000 = 1/6$

$P(B) = 2000/6000 = 1/3$

$P(C) = 3000/6000 = 1/2$

if x is the event that produce defective articles then

$P(x/A) = 1\% = 0.01$

$P(x/B) = 2\% = 0.02$

$P(x/C) = 3\% = 0.03$

$P(A/x) = ?$

$P(B/x) = ?$

$P(C/x) = ?$

Now we have,

$P(x) = P(A \cap x) + P(B \cap x) + P(C \cap x).$

$$= P(A).\,P(x/A) + P(B).\,P(x/B) + P(C).\,P(x/C)$$

$$= 1/6 \times 0.01 + 1/3 \times 0.02 + 1/2 \times 0.03$$

$$= 0.0233$$

So. i) $P(A/x) = \frac{P(A \cap x)}{P(x)} = \frac{P(A).\,P(x/A)}{P(x)}$

$$= \frac{1/6 \times 0.01}{0.0233}$$

$$= 0.0715$$

ii) $P(B/x) = \frac{P(B \cap x)}{P(x)} = \frac{P(B).\,P(x/B)}{P(x)}$

$$= \frac{1/3 \times 0.02}{0.0233}$$

$$= 0.2861$$

iii) $P(C/x) = \frac{P(C \cap x)}{P(x)} = \frac{P(C).\,P(x/C)}{P(x)}$

$$= \frac{1/2 \times 0.03}{0.0233}$$

$$= 0.6438$$

Hence, the probablity that the article is produced by machine A, machine B & machine C are 0.0715, 0.2861 & 0.6438 respectively.

073-11-10-3

chapter :-6 contd (Interference concerning mean).

6.5) One way ANOVA

ANOVA is statistical method for determining the existance of difference among several population mean. In fact treatments relative to the variance within treatment and hence the name analysis of variance (ANOVA).

It involves the statistical model either of data sampled from more than two populations or of data from experiments in which more than two treatment have been used.

It is a powerful statistical tool for the test of significance of homogenity of several means. It provides the comparison of the two estimates of the population variance using Fisher's F test, which is given by,

$$F = \frac{\text{variance between treatment}}{\text{variance within treatment}}$$

$$= \frac{\text{treatment mean square}}{\text{Error mean square}}$$

$$= \frac{MSTr}{MSE}$$

For testing the null hypothesis

H_0: the population mean μ_i ($i = 1, 2, 3, \cdots t$) are equal

against H_1: At least one μ_i differ from the other

The value of F should be $\simeq 1$ for the null hypothesis to be true & becomes large if μ_i differ significantly.

Thus, H_0 is rejected if $F \geq F_{\alpha(\nu_1, \nu_2)}$ at α level of significance.

$F_{\alpha(\nu_1, \nu_2)}$ $\nu_1 = t-1$ & $\nu_2 = n-t$

step 5 * calculation & decision

Treatments	observation (Replication)	sums	means
T_1	$y_{11}\ y_{12} \cdots y_{1j} \cdots y_{1r}$	$y_{1\cdot}$	$\bar{y}_1$
T_2	$y_{21}\ y_{22} \cdots y_{2j} \cdots y_{2r}$	$y_{2\cdot}$	$\bar{y}_2$
$\vdots$	$\cdots$	$\cdots$	$\vdots$
T_i	$y_{i1}\ y_{i2} \cdots \boxed{y_{ij}} \cdots y_{ir}$	$y_{i\cdot}$	$\bar{y}_i$
$\vdots$	$\cdots$	$\cdots$	$\vdots$
T_t	$y_{t1}\ y_{t2} \cdots y_{tj} \cdots y_{tr}$	$y_{t\cdot}$	$\bar{y}_t$

Grandsum = $y_{\cdot\cdot}$ Grandmean = $\bar{\bar{y}} = \frac{y_{\cdot\cdot}}{rt}$

(n = no. of observation)

* steps for calculation

i) correction factor (c) $= \frac{y_{\cdot\cdot}^2}{rt} = \frac{(\text{Grandsum})^2}{\text{no. of observations}}$

ii) sum of square of treatment (SSTr) $= \boxed{\frac{1}{r}\sum_{i=1}^{t} y_{i\cdot}^2 - c}$ ⇒ for equal

$= \boxed{\sum_{i=1}^{t}\left(\frac{y_{i\cdot}^2}{r_i}\right) - c}$ ⇒ for unequal

iii) sum of square of total (SST) $= \sum_{i}^{t}\sum_{j}^{r} y_{ij}^2 - c$

iv) sum of square of error (SSE) = SST − SSTr

v) mean squares,

Treatment mean square (MSTr) $= \frac{SSTr}{t-1}$

Error mean square (MSE) $= \frac{SSE}{n-t}$

Finally we form the ANOVA table.

008 Bhadra.

8) The output of three varities of wheat each grown on 4 plots of land is given below. Analysis the data and set up an ANOVA table state if the variety difference are significant at $\alpha = 0.05$ level.

varieties of wheat	yield tones / hectre			
A	6	7	3	8
B	5	5	3	7
C	5	4	3	4

Soln We set up the following hypothesis.

1) Null hypothesis

$H_0: \mu_1 = \mu_2 = \mu_3$ (There is no difference between the varieties of wheat)

2) Alternate hypothesis

$H_1: \mu_1 \neq \mu_2 \neq \mu_3$ (At least one differ from the other)

3) level of significance

$\alpha = 0.05$

4) Test statistics

$$F = \frac{MSTr}{MSE} = \frac{SSTr}{SSE} \times \frac{n-t}{t-1}$$

5) criteria

We reject H_0 if $F \geq F_{0.05}(2, 9) = 4.2565$

6) calculation & decision

variety of wheat	production				sum	mean
A	6	7	3	8	24	6
B	5	5	3	7	20	5
C	5	4	3	4	16	4

Grand sum = 60 , Grandmean = 5

i) correlation factor (c) = $\frac{(\text{Grandsum})^2}{\text{no. of observations}}$

$= \frac{(60)^2}{12} = 300$

ii) sum of square of treatment (SSTr) $= \frac{1}{4} \times 1232 - 300$

$= 8$

iii) sum of square of total (SST) $= 332 - 300$

$= 32$

iv) sum of square of error (SSE) = SST − SSTr

$= 32 - 8$

$= 24$

v) mean squares,

Treatment mean square (MSTr) $= \frac{SSTr}{t-1} = \frac{8}{3-1} = \frac{8}{2} = 4$

Error mean square (MSE) $= \frac{SSE}{n-t} = \frac{24}{12-3} = \frac{24}{9} = 2.66$

F ratio $= \frac{MSTr}{MSE} = \frac{4}{8/3} = 1.5$

one-way ANOVA table

sources of variation	degree of freedom	sum of square	mean square	F
varieties	$t-1 = 3-1 = 2$	$SST_r = 8$	$MST_r = 4$	1.5
error	$n-t = 12-3 = 9$	$SSE = 24$	$MSE = 2.667$	
Total	$n-1 = 11$	$SST = 32$		

Since $F_{0.05(2,9)} = 4.2565 > 1.5$ (calculated value of F). So there is no significance difference in yield of three varieties of wheat. The slight difference in sum & mean of the production is may be due to inherent characteristics.

72 Kartik

Q) In the investigation of a citizen's comittee complaint about the availability of fire protection within the country, the distance in miles to the nearest fire station was measured for each of five randomly selected residences in each of four areas.

$\nu_1 = t-1 \quad \nu_2 = n-t$

sol^n We set up the following hypothesis,

1) Null hypothesis

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

2) Alternate hypothes

$H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$

3) level of significance

$\alpha = 0.05$

4) Test statistics

$$F = \frac{MSTr}{MSE}$$

5) criteria

We reject H_0 if $F \geq F_{\alpha(\nu_1,\nu_2)} = F_{0.05(3,16)} = 3.24$

6) calculation & decision

Area	distance					sum	mean
1	7	5	5	6	8	31	6.2
2	1	4	3	4	5	17	3.4
3	7	9	8	7	8	39	7.8
4	4	6	3	7	5	25	5

Grand sum = 112 , Grand mean = 5.6

i) correlation factor (c) = $\frac{(\text{Grand sum})^2}{\text{no. of observation}}$

$$= \frac{(112)^2}{20} = 627.2$$

ii) sum of square of treatment (SSTr) $= \frac{1}{5} \times [3396] - 627.2$
$= 52$

iii) sum of square of total (SST) $= 708 - 627.2$
$= 80.8$

iv) sum of square of error (SSE) = SST − SSTr
$= 80.8 - 52$
$= 28.2$

v) mean square,

Treatment mean square (MSTr) $= \frac{SSTr}{t-1} = \frac{52}{3} = 17.33$

Error mean square (MSE) $= \frac{SSE}{n-t} = \frac{28.2}{16} = 12.2$

$\therefore$ F ratio $= \frac{MSTr}{MSE} = \frac{17.33}{12.2} = 1.42$

one way ANOVA table.

sources of variation	degree of freedom	sum of square	mean square	F
Area	$t-1 = 3$	$SSTr = 52$	$MSTr = 17.33$	1.42
error	$n-t = 16$	$SSE = 28.2$	$MSE = 12.2$	
Total.	$n-1 = 19$	$SST = 80.2$		

Since, $F < 3.24$, we accept H_0,
Hence, these data doesnot provide sufficient evidence to indicate a difference in mean distance for the four areas.